Titolo tesi: I numeri primi e l'Ipotesi di Riemann.

Autore: Zero87.

Versione senza frontespizio per matematicamente.it.

Archimedes will be remembered when Aeschylus is forgotten, because languages die and mathematical ideas do not. "Immortality" may be a silly word, but probably a mathematician has the best chance of whatever it may mean.

Hardy, "A mathematician's Apology" (cap. 8)

(Archimede sarà ricordato mentre Eschilo dimenticato, poiché i linguaggi muoiono ma le idee matematiche no. "Immortalità", potrebbe essere una parola stupida, ma probabilmente un matematico ha le migliori possibilità per avvicinarsi ad essa.)

INDICE

ABSTRACT

The purpose of this dissertation is to give an overview on the Riemann Hypothesis.
The Riemann Hypothesis – RH for short – is a question the German mathematician put about the distribution of the zeroes of an entire function derived from the analytic continuation of his zeta function, usually denoted by the Greek letter ζ.
Thus the basic reference of this dissertation is Riemann's original paper – *On the Number of Primes Less Than a Given Magnitude* (published in 1859) – where the connection between the zeta function and Number Theory is introduced and discussed.
We will start recalling some important results of mathematical Analysis, like sequences and series. We will extend them from the real line to the complex plane. In fact several theorems of Complex Analysis intervene in Riemann's paper, although without an explicit proof. One of the most important is Cauchy's residue theorem often used in order to compute improper integrals of real functions.
We also illustrate various basic concepts of Number Theory about prime numbers, arithmetical functions and Dirichlet series.
The main object of Riemann paper is the ζ function, that is the analytic continuation of the generalized harmonic series in the area $Re(s) > 1$ of the complex plane. Actually he defines the zeta function in the whole complex plane by an integral representation and two proofs of its functional equation (one of them is just based on Cauchy's residue theorem).
Then the German mathematician defines the xi function (denoted by Greek letter ξ) which is entire and – as we will see later – shares several noteworthy characteristics of the zeta function.
The original Riemann Hypothesis is about the zeroes of the ξ function but it was later adapted to the zeta function, so that his usual formulation deals with ζ.
The RH, in fact, says

<<All nontrivial zeta function's zeroes have real part one half.>>

The importance of RH is all about the connection between some properties of these two functions and the Number Theory.
However the emphasis of Riemann paper is not only on the Hypothesis. In his pages, in fact, we also find, for instance, an exact representation of Euler's quotient formula of primes (like the title itself of the paper suggests) obtained by the use of Möbius and Fourier inversion.

The history of the Riemann Hypothesis is interesting too. In fact this problem was raised, as said, by Riemann in the middle of the Nineteenth century, but it is still without a proof (prove to rejecting it). Several great mathematicians have attempted to solve it but unsuccessfully. During the Twentieth century the interest in this problem increased, as Hilbert inserted it in his famous list of twenty three problems ([21]) in the Paris First International Congress of Mathematicians in 1900. Therefore other attempts to give a proof of the Hypothesis or to

approach the problem from different perspectives were developed. For instance, Hardy clarified in 1914 part of the conjecture and showed the existence of infinitely many nontrivial zeroes of the zeta function in the so calling *critical line* (the line of the complex plane consisting of the points with real part one half). During the Twentieth century, other proofs similar to Hardy's were given but the problem is still far from a final solution.

INTRODUZIONE

Viaggiava lui.
Ed ogni volta finiva in un posto diverso: nel centro di Londra, su un treno in mezzo alla campagna, su una montagna così alta che la neve ti arrivava alla pancia, nella chiesa più grande del mondo, a contare le colonne e guardare in faccia i crocefissi. […] *Non c'era mai sceso da quella nave, mai sceso, proprio mai, non era una palla, era tutto vero. Mai sceso. Eppure era come se le avesse viste, tutte quelle cose. Novecento era uno che se tu gli dicevi "Una volta son stato a Parigi", lui ti chiedeva se avevi visto i giardini tal dei tali, e se avevi mangiato in quel dato posto, sapeva tutto, ti diceva "Quello che a me piace, laggiù, è aspettare il tramonto andando avanti e indietro sul Pont Neuf, e quando passano le chiatte, fermarmi e guardarle da sopra, e salutare con la mano".*
"Novecento, ci sei mai stato a Parigi, tu?"
"No."

Queste sono le parole di Max, trombettista del *Virginian* nel romanzo di Baricco *Novecento* ([4]). Rivediamo in esse il sensato stupore dell'uomo nel conoscere la storia del suo amico che, da quando è nato, non è mai sceso da quella nave nella quale era cresciuto e suonava il piano nell'orchestra di bordo. Eppure conosceva accuratamente il mondo esterno e, come apprendiamo in un altro passo, riusciva a leggere la gente ed a cogliere i sapori e gli odori delle loro terre dagli sguardi e dalle loro parole.

E' questo il mio obiettivo, proporre un lungo viaggio – più o meno difficile a seconda delle capacità di chi è intenzionato a seguirmi – lungo uno dei più importanti problemi aperti all'interno della matematica: l'ipotesi di Riemann.
Sono il primo a dire che questo non è un viaggio per tutti e, come in ogni impresa, c'è *chi può farla e chi non può*. Non si tratta di razzismo intellettuale o altri idealismi che, francamente, sono ben lieto di tenere lontano dagli intenti di questo lavoro.
Per affrontare la comprensione di questo problema – inserito nella lista dei sette problemi del millennio ([22]) – occorre avere un minimo di conoscenze di base, diciamo di Analisi I.
Il resto ce lo metto io cercando, di volta in volta, di fornire i mezzi necessari alla comprensione delle tematiche che reca con sé questo grande interrogativo matematico. Ci saranno sezioni dedicate a richiami di Analisi Matematica, alla Teoria dei Numeri... per poi affrontare l'ipotesi di Riemann nella sua interezza. Qui la scalata potrebbe rivelarsi troppo difficile ma spero di riuscire a farla affrontare a chiunque soddisfi i requisiti ed abbia voglia giungere a questa idea che ha tormentato le migliori menti matematiche degli ultimi due secoli.
Gli ultimi capitoli tratteranno risultati che andranno al di là dell'ipotesi in sé: si vedranno delle conseguenze della stessa oltre a formule successive trovate a partire dal lavoro di Riemann (von Mangoldt, Siegel, ...). Molti di questi risultati, proprio perché aggiuntivi (ma anche a causa della loro complessità), saranno, generalmente, non dimostrati.

Le parti finali di questa tesi conterranno delle sezioni più tecniche nelle quali verranno esposti altri risultati fondamentali della Teoria Analitica dei Numeri come il Teorema dei Numeri Primi con relativa dimostrazione.

1. RICHIAMI DI ANALISI MATEMATICA I

In questa sezione verranno richiamati tre concetti fondamentali riguardanti la "prima" analisi matematica con cui si ha a che fare quando si intraprende lo studio della materia a livello universitario.
Parleremo, dunque, di successioni, serie e formula di Taylor, supponendo come acquisiti altri passi intellettivi come gli integrali e gli studi di funzione. La scelta di fare il punto su successioni e serie è motivata principalmente dal fatto che la stessa ipotesi di Riemann riguarda gli *zeri* di una funzione che non è altro che l'estensione – nel campo complesso – di una serie.

1.1 SUCCESSIONI

1.1.1 Successioni in $\mathbb{R}$

Una successione in un insieme A è una funzione f che ad ogni numero naturale – o comunque ad ogni numero naturale sufficientemente grande – associa un elemento di A. Si scrive in genere $f(0) \equiv a_0, f(1) \equiv a_1, \dots, f(n) \equiv a_n$ e così via. Per questo, con un leggero abuso di scrittura, indicheremo con $(a_n)_{n\in I}$ la nostra successione. Ripetiamo che si suppone che $I \subseteq \mathbb{N}$ coincida con tutto $\mathbb{N}$ o con i naturali maggiori o uguali di un $n_0 \in \mathbb{N}$ prefissato.
Saremo principalmente interessati al caso in cui A coincide con l'insieme $\mathbb{R}$ dei reali, ma ci dedicheremo anche al caso in cui A è l'insieme dei complessi.
Generalmente introdurremo una successione precisando un valore generico di $f(n)$. Per esempio la scrittura

$$\left(\frac{n}{n+1}\right)_{n\in\mathbb{N}}$$

rappresenta la successione infinita che assume i valori $a_0 = 0, a_1 = \frac{1}{2}, a_2 = \frac{2}{3}, \dots, a_n = \frac{n}{n+1}, \dots$ che si ottengono semplicemente sostituendo il numero naturale desiderato all'interno della funzione.
Ma una successione può essere anche definita per *ricorsione*, precisandone gli elementi iniziali e poi fissando, per ogni n, una legge che determina l' $(n+1)-$esimo termine mediante i precedenti.
Uno degli esempi più famosi è la successione dei numeri di Fibonacci che possiamo definire, proprio per ricorsione, nel modo seguente:

$$\begin{cases} a_0 = 1 \\ a_1 = 1 \\ a_{n+1} = a_n + a_{n-1} \end{cases}$$

In questo caso, a partire dai primi due termini, possiamo dunque calcolare gli altri mediante la legge espressa per ricorsione.

$$a_2 = a_1 + a_0 = 1 + 1 = 2$$
$$a_3 = a_2 + a_1 = 1 + 2 = 3$$
$$a_4 = a_3 + a_2 = 3 + 2 = 5$$
$$a_5 = a_4 + a_3 = 5 + 3 = 8$$

e così via. Tuttavia la complessità di questa scrittura – apparentemente intuitiva – cresce molto all'aumentare dell'indice dell'elemento che vogliamo calcolare. Supponiamo, infatti, di voler sapere quanto vale a_{100}: ci arriveremo andando avanti con il metodo appena visto dopo aver calcolato i primi 99 elementi. Ma questo non è un problema dovuto alla rappresentazione della sequenza: a meno che non si abbia a che fare con successioni definite mediante una formula che, al variare dell'indice, definisce direttamente il corrispettivo valore senza passaggi intermedi (tipo la già citata $\left(\frac{n}{n+1}\right)_{n\in\mathbb{N}}$), in tutti gli altri casi ricorre lo stesso problema.

Definiamo, ora, il limite di una successione in $\mathbb{R}$.

Diremo che $L \in \mathbb{R}$ è il limite della $(a_n)_{n\in\mathbb{N}}$ e scriveremo

$$\lim_{n\to\infty} a_n = L,$$

quando per ogni reale $\varepsilon > 0, \exists \bar{n} \in \mathbb{N}$ tale che $a_n \in [L - \varepsilon, L + \varepsilon], \forall n > \bar{n}, n \in \mathbb{N}$.

In questo caso si dice che la successione $(a_n)_{n\in\mathbb{N}}$ converge a L e si scrive anche $a_n \to L$.

Si utilizzano indifferentemente le due scritture

$$\lim_{n\to+\infty} a_n = L = \lim_{n\to\infty} a_n = L,$$

omettendo il segno dell'infinito, poiché l'indice n è naturale dunque può tendere soltanto all'infinito di segno positivo.

Ora, accanto alle successioni convergenti, ne esistono anche altre che non convergono e, di queste, alcune sono chiamate divergenti. Diremo che una successione di reali è divergente a $+\infty$ (o a $-\infty$) e scriveremo

$$\lim_{n\to\infty} a_n = +\infty, \qquad \left(\text{o} \lim_{n\to\infty} a_n = -\infty\right),$$

se per ogni intero $M > 0, \exists \bar{n} \in \mathbb{N}$ tale che $a_n > M$ (o $a_n < -M$), $\forall n > \bar{n}, n \in \mathbb{N}$.

Un esempio di sequenza divergente (a $+\infty$) è la seguente

$$\left(\frac{1}{n+1} + n\right)_{n\in\mathbb{N}}$$

per la quale si può facilmente verificare che $a_n \to +\infty$ per $n \to +\infty$.

Esiste un terzo caso di successioni che non ammettono limite e, dunque, non sono né convergenti né divergenti. Di questo tipo di sequenza, un esempio può essere

$$(\cos(\pi n))_{n\in\mathbb{N}},$$

che non ammette limite in quanto oscilla tra i valori -1 e 1 senza tendere definitivamente a nessuno di loro.

Consideriamo ora la successione definita ponendo

$$a_n = \frac{1}{n}, \qquad \forall n \in \mathbb{N}\backslash\{0\}$$

quindi $\left(\frac{1}{n}\right)_{n\in\mathbb{N}\setminus\{0\}}$: possiamo notare che $\left(\frac{1}{n}\right)_{n\in\mathbb{N}\setminus\{0\}} \subseteq [0,1]$.
In questo, e in altri casi analoghi, la successione si dice *limitata* proprio perché i suoi valori assunti sono compresi in un intervallo limitato di $\mathbb{R}$ ([19], §3.1). In caso contrario, la successione viene detta illimitata. Un esempio di sequenza illimitata, invece, è quella già esaminata in precedenza

$$a_n = \frac{1}{n+1} + n, \qquad \forall n \in \mathbb{N},$$

nella quale, come già detto, $a_n \to +\infty$ per $n \to +\infty$. Abbiamo considerato esempi semplici, ma quanto detto vale per qualsiasi tipo di successione.
Una successione è crescente se $a_{n+1} \geq a_n \ \forall n$, mentre è decrescente se $a_{n+1} \leq a_n \ \forall n$. In entrambi i casi la sequenza è detta anche monotòna (crescente o decrescente). Per le successioni monotòne vale un interessante risultato ([2], §10.3).

Teorema
Una successione monotòna converge se e solo se è limitata.

Un esempio utile di successione monotòna (in questo caso crescente) è il seguente

$$a_n = \left(1 + \frac{1}{n}\right)^n, \qquad \forall n \in \mathbb{N}\setminus\{0\}.$$

Si prova che questa successione è convergente (vedi, ad es. [19], §3.30) e il suo limite è detto numero di Nepero

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n = e = 2{,}71828 \ldots$$

Tuttavia per una stima accurata del numero di Nepero non si utilizza questa successione ma si preferiscono metodi alternativi: uno di essi, legato alla formula di Taylor, sarà esaminato nell'ultimo paragrafo di questa sezione.

1.1.2 Risultati sulle successioni

In questo paragrafo tratteremo altri risultati importanti sulle successioni in $\mathbb{R}$.
Molti di essi sono piuttosto teorici e valgono, identici, anche per successioni a valori complessi, perciò non saranno nuovamente richiamati in seguito, nella sezione dedicata all'analisi complessa.

Teorema ([19], §3.2)
Sia $(a_n)_{n\in\mathbb{N}}$ una successione a valori in $\mathbb{R}$.
- Se $a, a' \in \mathbb{R}$ sono tali che $a_n \to a$ e $a_n \to a'$ allora $a = a'$.
- Se a è un punto limite di $A \subseteq \mathbb{R}$ allora $\exists \{a_n\} \subseteq A$ t.c. $a_n \to a$.

Il primo risultato in molti testi è trattato in maniera indipendente come "teorema di unicità del limite" e ci dice proprio che una successione convergente non può avere due limiti differenti.

Il secondo, invece, è molto più importante di quello che sembra e riguarda i punti limite degli insiemi, siano essi massimi, minimi, estremi superiori,... Questo risultato, infatti, ci dice che se abbiamo un punto limite di un insieme – sia esso un massimo o un minimo – allora esiste una successione contenuta in esso che converge in quel punto. Viene utilizzato in molti teoremi di Analisi I (e non solo).
Questo risultato varrà anche per successioni a valori complessi.

Teorema
Siano a_n, b_n successioni a valori reali tali che $a_n \to a \in \mathbb{R}$ e $b_n \to b \in \mathbb{R}$. Siano inoltre $c, d \in \mathbb{R}$. Allora

- $\lim_{n\to\infty}(a_n \pm b_n) = \lim_{n\to\infty} a_n \pm \lim_{n\to\infty} b_n = a \pm b$.
- $\lim_{n\to\infty} c \cdot a_n = c \cdot \lim_{n\to\infty} a_n = c \cdot a$.
- $\lim_{n\to\infty} 1/a_n = 1/a$ (assumendo ovviamente $a_n \neq 0, \forall n$ abbastanza grande: si noti che non è detto che $a \neq 0$, ma anche in questo caso il risultato resta valido assumendo $1/0 = \infty$ con il segno opportuno).

Questo teorema può riassumersi agilmente nella seguente scrittura:

$$\lim_{n\to\infty}(c \cdot a_n + d \cdot b_n) = c \cdot \lim_{n\to\infty} a_n + d \cdot \lim_{n\to\infty} b_n = c \cdot a + d \cdot b.$$

1.1.3 Successioni di Cauchy e completezza

Una successione $(a_n)_{n\in\mathbb{N}}$ è detta successione di Cauchy se $\forall \varepsilon > 0$, esiste un intero positivo N tale che $|a_n - a_m| < \varepsilon, \forall n, m \in \mathbb{N}$ tale che $n, m > N$.

Teorema
In $\mathbb{R}$ (ma anche in $\mathbb{R}^k, k \geq 1$) tutte le successioni di Cauchy convergono. In generale le successioni convergenti in $\mathbb{R}$ (e $\mathbb{R}^k$) sono tutte e sole quelle di Cauchy.

A sottolineare questa proprietà, diremo che $\mathbb{R}$ è *completo*. In questa sezione non tratteremo a fondo le successioni di Cauchy e non parleremo di sottosuccessioni poiché questi argomenti non interessano direttamente l'obiettivo fissato, ovvero l'ipotesi di Riemann. Rimandiamo però chi desidera approfondimenti alla lettura di un qualsiasi testo di analisi I, consigliando in particolare *Principles of Mathematical Anaysis*, di W. Rudin ([19]), da §3.5 a §3.19.

1.1.4 Successioni particolari

In matematica ci sono varie successioni meritevoli di attenzione. Una di esse è la già citata sequenza dei numeri di Fibonacci, recentemente portata alla ribalta dal best seller di Dan

Brown "Il codice da Vinci". Tuttavia, letteratura a parte, ve ne sono altre di particolare riguardo. Ne citiamo alcune tra le più importanti.

1. $a_n = \left(1 + \frac{1}{n}\right)^n$, definita per $n \in \mathbb{N}\backslash\{0\}$. Come già osservato essa è la successione che ha come limite il numero e. In ogni testo di analisi I è possibile trovare la dimostrazione del fatto che essa è monotòna crescente e limitata.
   In generale $\left(1 + \frac{1}{n}\right)^{\alpha n} \to e^{\alpha}, \forall \alpha \in \mathbb{R}$.
2. Se $x \in \mathbb{R}$ è t.c. $|x| < 1$, allora $a_n = x^n \to 0$.
3. Consideriamo la successione definita da $a_n = \frac{1}{n^k}$, per $n \neq 0$. Per $k > 0$ è limitata e $a_n \to 0$. Dalle proprietà delle potenze è semplice verificare che per $k < 0$ la stessa successione è divergente poiché $\frac{1}{n^k} = n^{-k}$ e $-k > 0$. Ovviamente per $k = 0$ la successione si riconduce ad un'infinità di termini costanti che assumono tutti valore 1 (e diverge).
4. Se $p > 0$, allora $a_n = \sqrt[n]{p} \to 1$ (con $n \neq 0$). Per $p = 0$ la radice $n-$esima non ha senso mentre assumendo $p < 0$ si entra nel campo dei numeri complessi così che rimandiamo il discorso al capitolo relativo.
5. La successione definita come $a_n = \sqrt[n]{n}$, definita per $n \in \mathbb{N}\backslash\{0\}$, tende anch'essa a 1.

Le ultime tre successioni sono utilizzate come modello in casi più complessi come altre successioni oppure successioni di funzioni.

1.2 SERIE

1.2.1 Introduzione alle serie: sommatoria e prodotti infiniti

Prima di passare alla trattazione delle serie, sembra doveroso richiamare brevemente il concetto di sommatoria ed alcune sue proprietà.
La sommatoria è una particolare scrittura matematica utilizzata per abbreviare una somma di elementi di particolari insiemi: in questi casi si rivela essere molto utile poiché riassume in forma compatta una rappresentazione anche piuttosto lunga e dispersiva.
Abbiamo, dunque, una rappresentazione di questo genere:

$$\sum_{i=n}^{m} a_i = a_n + a_{n+1} + \ldots + a_{m-1} + a_m$$

In essa

- i è l'indice di riferimento al quale applicare l'operazione di somma
- a_i rappresenta gli elementi che si intendono sommare al variare dell'indice i
- n e m sono gli estremi (naturali o interi) dell'intervallo rispetto al quale varia l'indice della sommatoria.

La scrittura appena vista si legge "sommatoria per i che varia da n a m di a_i"; ovviamente per avere senso si suppone che $n < m$. Un modo stilistico alternativo di indicare l'intervallo della sommatoria è quello di disporre gli indici a lato del simbolo: $\sum_{i=n}^{m} a_i$.
In quella scrittura, gli elementi indicati con a_i sono l'argomento della sommatoria e possono essere un qualsiasi tipo di oggetti matematici tra i quali è possibile attuare un'operazione di somma. L'intervallo può essere tra due numeri naturali o interi e non è necessariamente finito.
Vediamo di fare qualche esempio.

$$\sum_{n=2}^{5} n^2 = 2^2 + 3^2 + 4^2 + 5^2 = 4 + 9 + 16 + 25 = 54$$

In questo caso l'indice della sommatoria varia tra un intervallo di numeri naturali. Passiamo a

$$\sum_{n=1}^{4} x^n = x + x^2 + x^3 + x^4$$

Questa volta la sommatoria restituisce il polinomio $x + x^2 + x^3 + x^4$. Consideriamo ora

$$\sum_{i=n}^{m} \cos(i\pi) = \cos(n\pi) + \cos\big((n+1)\pi\big) + \ldots + \cos\big((m-1)\pi\big) + \cos(m\pi).$$

Questa volta la notazione è più generale e rappresenta una somma di coseni.
Si introducono anche sommatorie riferite a infiniti addendi (dunque ad un insieme infinito di indici). Formalmente, la sommatoria infinita, per esempio infiniti addendi reali a_i per i che varia tra i naturali maggiori o uguali di un certo n, è intesa nel modo seguente

$$\sum_{i=n}^{\infty} a_i = \lim_{m\to\infty} \sum_{i=n}^{m} a_i.$$

Un esempio può essere il seguente

$$\sum_{n=0}^{\infty} \frac{1}{n^x} = 1 + \frac{1}{2^x} + \frac{1}{3^x} + \ldots + \frac{1}{m^x} + \ldots \qquad x \in \mathbb{R}.$$

Questa sommatoria infinita che restituisce la somma dei reciproci dei naturali elevati a potenza non è un esempio preso a caso: nelle sezioni future vedremo che è un argomento centrale di questa tesi.
Quindi, l'operazione di sommatoria, non è altro che una somma di elementi indicizzati i cui indici variano all'interno di un intervallo di numeri naturali o interi (eventualmente infinito).
Generalmente, come estremi degli intervalli sui quali si opera la somma si utilizzano gli indici m, n mentre come indici della sommatoria sono impiegate le lettere i, j, k, l (ma anche n, m se l'intervallo di somma non è limitato e quindi non compaiono m e n come estremi).
Accanto al modo appena descritto di indicare una sommatoria ve ne sono altri di uso comune. Per esempio

$$\sum_{i\in I} a_i$$

rappresenta la somma tra gli elementi che hanno l'indice che varia all'interno di uno specifico insieme (o sottoinsieme) I di interi, mentre

$$\sum_{n\leq i\leq m} a_i \equiv \sum_{i=n}^{m} a_i$$

non è altro che un modo diverso per indicare la notazione standard.
Un prodotto $\prod$ ha una definizione formale analoga alla sommatoria con la differenza che, con questa scrittura, abbreviamo il prodotto tra elementi (indicizzati) di uno stesso insieme

$$\prod_{i=n}^{m} a_i = a_n \cdot a_{n+1} \cdot \ldots \cdot a_{m-1} \cdot a_m$$

In essa valgono tutte le rappresentazioni già viste nella sommatoria. Un utilizzo tipico della produttoria è nella rappresentazione del calcolo del fattoriale di un naturale n

$$n! = \prod_{k=1}^{n} k = 1 \cdot 2 \cdot \ldots \cdot (n-1) \cdot n$$

Analogamente alla sommatoria, la produttoria può non essere finita. In questo caso si parla anche di prodotto infinito e l'estensione naturale a questo caso è la seguente

$$\prod_{k=n}^{\infty} a_k = \lim_{m \to +\infty} \prod_{k=n}^{m} a_k$$

ammesso che un tale limite esista. Infatti, esempi come

$$\prod_{k=n}^{\infty} (-k)$$

non ammettono limite in quanto il valore assunto dal prodotto cambia continuamente segno al crescere dell'indice k alternando valori positivi a negativi senza né convergere né divergere.

1.2.2 Proprietà della sommatoria e dei prodotti (infiniti)

Sommatoria e prodotti di più addendi o fattori mantengono le proprietà usuali dell'addizione e della moltiplicazione (almeno finché restano finiti, perché altrimenti sorgono diversi problemi che vanno trattati più nello specifico caso con cui si ha a che fare). Inoltre ci sono anche altre proprietà che li collegano a funzioni particolari come l'esponenziale e il logaritmo. In questo paragrafo ne citiamo alcune tra le più importanti: altre – se necessario – saranno adeguatamente richiamate a tempo debito. Tutte queste proprietà si possono trovare in un qualsiasi testo di analisi I o calcolo o, più semplicemente, in ([11]).
Indicando con n e m dei qualsiasi valori (interi o naturali) e $n < m$, valgono le seguenti proprietà.

1. Se c è una costante di qualsiasi tipo (intera, reale,…), allora
$$\sum_{i=n}^{m} (c \cdot a_i) = c \cdot \sum_{i=n}^{m} a_i.$$
2. Per a_i e b_i argomenti di sommatorie con eguali indici, si ha
$$\sum_{i=n}^{m} a_i \pm \sum_{i=n}^{m} b_i = \sum_{i=n}^{m} (a_i \pm b_i).$$
3. Per k costante intera, vale

$$\sum_{i=n}^{m} a_i = \sum_{i=n+k}^{m+k} a_{i-k},$$

operazione che prende anche il nome di traslazione degli indici della sommatoria.

4. Per k intero, $n \leq k \leq m$, vale

$$\sum_{i=n}^{m} a_i = \sum_{i\in[n,k]} a_i + \sum_{i\in(k,m]} a_i,$$

proprietà, spesso utilizzata per isolare uno o più indici della sommatoria stessa.

5. Per a_i e b_i argomenti di produttorie con eguali indici, si ha

$$\left(\prod_{i=n}^{m} a_i\right) \cdot \left(\prod_{i=n}^{m} b_i\right) = \prod_{i=n}^{m} a_i b_i,$$

proprietà largamente utilizzata che trae il suo fondamento dalla proprietà commutativa della moltiplicazione. Vale l'analogo

$$\left(\prod_{i=n}^{m} a_i\right) \Big/ \left(\prod_{i=n}^{m} b_i\right) = \prod_{i=n}^{m} a_i / b_i.$$

6. Considerando le proprietà degli esponenziali

$$x^{\sum_{i=n}^{m} a_i} = \prod_{i=n}^{m} x^{a_i}.$$

7. Dalle proprietà del logaritmo

$$\sum_{i=n}^{m} \log(a_i) = \log\left(\prod_{i=n}^{m} a_i\right).$$

Tutte queste proprietà si possono anche estendere, con le dovute precauzioni, anche in casi infiniti (n o m o entrambi infiniti, ammesso che non si abbia a che fare con forme indeterminate). Tuttavia occorre fare qualche piccola osservazione in merito alle proprietà 6. e 7.. In esse, infatti, si dà per scontato che gli esponenziali e i logaritmi sono "reali", cioè funzioni reali di variabile reale e presuppongono l'iniettività delle funzioni in gioco: vedremo, infatti, che nel caso complesso le funzioni esponenziali e il logaritmo non sono iniettive e che addirittura ad un valore nella loro immagine ne corrispondono infiniti nella retroimmagine.

1.2.3 Le serie

Data una successione $(a_n)_{n\in\mathbb{N}}$, ad esempio di numeri reali, possiamo considerare la somma progressiva di alcuni (o tutti) i suoi termini. In essa n e m indicano qui i relativi estremi

$$\sum_{i=n}^{m} a_i = a_n + a_{n+1} + \ldots + a_{m-1} + a_m$$

Essa è detta serie (finita) e non è altro che la somma dei termini di una successione: come tale, possiede tutte le proprietà appena viste sulle sommatorie.

Salvo indicazioni contrarie, con il termine "serie" si intende una somma infinita, cioè una somma nella quale almeno uno dei due estremi non è finito, quindi con infiniti addendi. Anzi, in generale si utilizza il termine "serie" proprio in luogo di "sommatoria infinita". In questo caso si può anche usare la seguente abbreviazione in linea a quanto detto per le sommatorie (con n intero)

$$\sum_{i=n}^{\infty} a_i \equiv \sum_{i \geq n} a_i.$$

Se la serie è infinita e l'intervallo inizia dallo zero (o dall'indice $n_0 \geq 1$, a seconda se *sia possibile definire o meno i termini minori di* n_0), ci sono altre scritture comunemente utilizzate nelle quali si omettono gli estremi dell'intervallo o si lascia l'indice

$$\sum_{n=0}^{\infty} a_n \equiv \sum a_n \equiv \sum_n a_n$$

Tuttavia, tralasciando le formalità stilistiche, ad ogni serie infinita della forma appena descritta si può associare una sequenza, detta *successione delle somme parziali* $\{S_n\}$ definita nel modo seguente per ogni n naturale

$$S_n = \sum_{i=0}^{n} a_i = a_0 + a_1 + \ldots + a_n$$

La differenza (per ora solo formale) tra la serie e la somma parziale viene detta somma residua

$$R_n = \sum_{i=0}^{\infty} a_n - S_n = \sum_{i=0}^{\infty} a_n - \sum_{i=0}^{n} a_n = \sum_{i=n+1}^{\infty} a_n$$

E' semplice osservare che

$$\lim_{n\to\infty} S_n = \lim_{n\to\infty} \sum_{i=0}^{n} a_i = \sum_{i=0}^{\infty} a_i$$

Da queste considerazioni possiamo convenire che la serie $\sum a_i$ converge se è convergente la sequenza delle somme parziali. In quel caso scriveremo

$$\sum_{n=0}^{\infty} a_n = s = \lim_{n\to\infty} S_n$$

nel quale $s \in \mathbb{R}$ è il valore a cui converge la serie. Analogamente, se la successione $\{S_n\}$ diverge, diremo che la serie è divergente.

Sovente, soprattutto in testi di analisi, si utilizza la seguente notazione

$$\sum a_n < \infty$$

per definire una serie convergente.

1.2.4 Risultati sulle serie

Per le serie ci sono molti risultati per ciò che concerne la convergenza; ne riporteremo alcuni. Tra questi ce ne sarà qualcuno che riguarderà delle serie particolari con cui avremo a che fare nelle sezioni successive.

Teorema
$\sum a_n$ converge se e solo se per ogni $\varepsilon > 0$ esiste un intero positivo N t.c.

$$\left|\sum_{k=n}^{m} a_k\right| \leq \varepsilon, \qquad \forall m, n \in \mathbb{N}, m \geq n \geq N.$$

Teorema (classico)
Se $\sum a_n$ converge allora $a_n \to 0$.

La condizione $a_n \to 0$, tuttavia, è necessaria ma non sufficiente a garantire la convergenza di una serie. Tuttavia, anche qui possiamo affermare che se $a_n \nrightarrow 0$ allora la serie non converge.

Esempio ([19], §3,28)
La serie

$$\sum_{n=1}^{\infty} \frac{1}{n^k}$$

è detta serie armonica generalizzata e converge per $k > 1$. Se $k = 1$ la serie è armonica (semplice) ed è divergente.

Teorema (criteri di confronto)
Siano N un intero positivo fissato e a_n, c_n, d_n successioni definite sui naturali a termini non negativi e N un intero positivo fissato.

- Sia b è una costante reale positiva fissata. Se vale $a_n \leq b \cdot c_n$ per ogni $n \geq N$ allora se $\sum c_n$ converge anche $\sum a_n$ converge.
- Se $a_n \geq d_n \geq 0$ per $n \geq N$, se $\sum d_n$ diverge, diverge anche $\sum a_n$.

I criteri di confronto sono molto utili per gli studi delle serie. Per fare un esempio, supponiamo di voler vedere se converge o diverge la seguente serie

$$\sum_{n=1}^{\infty} \frac{n+1}{n}.$$

Possiamo confrontarla con una serie conosciuta

$$\sum_{n=1}^{\infty} \frac{n+1}{n} = \sum_{n=1}^{\infty} \left(1 + \frac{1}{n}\right) > \sum_{n=1}^{\infty} \frac{1}{n}$$

concludendo che è divergente, essendo divergente la serie armonica utilizzata per la maggiorazione.

Esempio ([19], §3,26)
Per $x \geq 0$ la serie

$$\sum_{n=0}^{\infty} x^n$$

è detta serie geometrica e converge per $0 \leq x < 1$ mentre diverge se $x > 1$. Inoltre, per $x \in [0,1)$

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$$

Teorema ([19], §3.27)
Supponiamo $a_1 \geq a_2 \geq \ldots \geq a_n \geq \ldots$. Allora la serie $\sum a_n$ converge se e solo se converge la serie

$$\sum_{k=0}^{\infty} 2^k a_{2^k} = a_1 + 2a_2 + 4a_4 + 8a_8 + \cdots$$

Esempio ([19], §3.29)
La serie

$$\sum_{n=2}^{\infty} \frac{1}{n(\log n)^p}$$

converge per $p > 1$ mentre diverge per $p \leq 1$. Da notare che, a causa del logaritmo e del quoziente, essa non può essere definita nei casi $n = 0$ e $n = 1$.

Teorema (comparazione del limite)
Siano $\sum a_n$ e $\sum b_n$ due serie distinte a termini positivi. Se

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 1$$

allora la serie $\sum a_n$ converge se e solo se converge $\sum b_n$.

Teorema (test delle radici)
Data la serie $\sum a_n$, sia $l = \lim_{n\to\infty} \sqrt[n]{|a_n|}$.

- Se $l < 1$ la serie converge.
- Se $l > 1$ la serie diverge.
- Se $l = 1$ non possiamo stabilire il carattere della serie.

Teorema (test della razionalità)
Data la serie $\sum a_n$, consideriamo la quantità

$$\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = l$$

- Se $l < 1$ la serie converge.
- Se $l \geq 1$ la serie diverge.

1.2.5 Convergenza assoluta

Per quanto riguarda le serie, fino ad ora si sono considerati soprattutto risultati riguardanti serie a termini positivi o, comunque, non negativi. Le serie, però, possono essere di qualsiasi tipo: a termini negativi, alternati o misti. Occorre quindi qualche criterio per studiare il carattere – cioè la convergenza o divergenza – di una serie più generale.
Generalmente, data una serie $\sum a_n$ ci si basa sulla serie dei valori assoluti $\sum |a_n|$ per trarre delle conclusioni anche a suo riguardo: infatti, ricordando la disuguaglianza triangolare

$$|a+b| \leq |a| + |b|, \qquad \forall a, b \in \mathbb{R}$$

ed estendendola al caso generale di n termini (vale anche per $n \to \infty$)

$$\left| \sum_{i=0}^{n} a_n \right| \leq \sum_{i=0}^{n} |a_n|$$

possiamo capire il carattere della serie generale a partire da quella dei termini in modulo. Inoltre $|a_n| \geq 0$ quindi per la serie dei valori assoluti valgono tutti i risultati visti nel paragrafo precedente su serie a termini non negativi.
Diremo che la serie $\sum a_n$ converge assolutamente se la serie $\sum |a_n|$ è convergente. Tuttavia se la serie di partenza è già a termini non negativi, convergenza semplice e assoluta sono la stessa cosa.

Teorema
Se $\sum_{n=0}^{\infty} a_n$ converge assolutamente allora converge.

Di questo risultato non vale il viceversa, cioè la convergenza semplice non implica necessariamente quella assoluta. Nel caso in cui una serie converge ma non assolutamente, essa viene anche detta condizionalmente convergente.

1.3 LA FORMULA DI TAYLOR

La formula di Taylor è un risultato fondamentale per lo studio locale di funzioni sufficientemente regolari. Essa consente di approssimare in maniera efficace una funzione – anche complicata – in un intorno servendosi solamente di determinati valori della stessa.
La formula di Taylor, infatti, approssima una qualsiasi funzione sufficientemente regolare con un polinomio, o meglio, con una serie di potenze la quale, a rigor di fatti, è essa stessa un polinomio anche se un po' particolare (se ne parlerà nella seguente sezione di richiami di Analisi Matematica II). Questa formula ci consente, quindi, di avere una conoscenza del comportamento di una funzione in un intorno con una precisione arbitraria, la quale è maggiore all'aumentare dei termini che consideriamo.

1.3.1 Introduzione

Consideriamo, per ora, una funzione f di classe C^1, $f: A \to \mathbb{R}$, con A intervallo aperto, e un punto $x_0 \in A$. Sappiamo che in un intorno di x_0 possiamo approssimare f con un polinomio di primo grado

$$f(x) \cong f(x_0) + f'(x_0)(x - x_0)$$

per x vicino a x_0. Questa è un'approssimazione *lineare* della funzione f. Con il simbolo $\cong$ intendiamo che la differenza tra il primo ed il secondo membro, che indichiamo con $R_1(x)$ (resto di ordine 1), tende a zero più rapidamente di $x - x_0$. Dunque:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + R_1(x),$$

$$\text{in cui} \qquad \lim_{x \to x_0} \frac{R_1(x)}{x - x_0} = 0.$$

Questa approssimazione, tuttavia, è molto debole ed è efficace proprio nelle immediate vicinanze del punto x_0. Se, ad esempio, consideriamo la funzione $f(x) = e^x$ in un intorno del punto $x_0 = 0$, abbiamo la seguente approssimazione

$$e^x \cong e^{x_0} + e^{x_0}(x - x_0) = e^0 + e^0(x - 0) = 1 + x$$

A questo punto usiamo la formula per calcolare il valore di $f(x)$ in determinati punti.

- Per $x = 0{,}1$, l'approssimazione ci dà come risultato $1 + 0{,}1 = 1{,}1$ ma, in realtà, il valore di $e^{0{,}1}$ è $e^{0{,}1} = 1{,}10517$ ...; commettiamo un errore di $0{,}00517$ ... tra il valore *vero* e quello approssimato con la formula vista in precedenza.
- Per $x = 0{,}2$, l'approssimazione ci dà come risultato $1 + 0{,}2 = 1{,}2$ ma, in realtà, il valore di $e^{0{,}2}$ è $e^{0{,}2} = 1{,}2214$ ...; si commette un errore di $0{,}0214$ ...
- Per $x = 0{,}5$, l'approssimazione dà come risultato $1 + 0{,}5 = 1{,}5$ ma, in realtà, il valore di $e^{0{,}5}$ è $e^{0{,}5} = 1{,}64872$ ...; l'errore è di $0{,}14872$ ...
- Per $x = 1$, l'approssimazione dà come risultato $1 + 1 = 2$ ma $e^1 = e = 2{,}71828$ ...; l'errore è $0{,}71828$ ...
- Per $x = 5$, l'approssimazione dà come risultato $1 + 5 = 6$ ma $e^6 = 148{,}413$ ...

Notiamo che, per valori sempre più lontani dal punto centrale, l'approssimazione è inefficace.
Supponiamo, dunque, di avere una funzione di classe C^n, per $n \geq 1$, ci si può domandare se sia possibile ottenere un miglior grado di approssimazione rispetto al caso $n = 1$, cioè se sia possibile scomporre $f(x)$ in un polinomio di grado n e resto $R_n(x)$ che tenda a zero più rapidamente di $(x - x_0)^n$. La risposta sta proprio nella formula di Taylor.

1.3.2 La formula di Taylor

Sia $f(x)$ una funzione di classe C^n in un intorno di x_0. Risulta

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + R_n(x),$$

$$\text{in cui} \qquad \lim_{x \to x_0} \frac{R_n(x)}{(x - x_0)^n} = 0.$$

In questa formula, per $k = 0$ si intende $f^{(0)}(x_0) = f(x_0)$ e $0! = 1$. Essa viene chiamata formula di Taylor, serie di Taylor (lo vedremo nella prossima sezione) o anche polinomio di Taylor. Il punto x_0 è il centro di questo sviluppo e il modo più comune per definirla è proprio "polinomio (o serie) di Taylor di f centrata in x_0 (di ordine n)". Se $x_0 = 0$ il polinomio è anche detto "polinomio (o serie) di McLaurin (di ordine n)".

Il teorema di Taylor si può anche rovesciare in un modo molto interessante ([2], §7.2).
Sia $P(x)$ un polinomio di grado $\leq n$ definito nel modo seguente in $x_0 \in [a, b]$

$$P(x) = \sum_{k=0}^{n} \frac{f^{(n)}(x_0)}{k!} (x - x_0)^k$$

con $f(x)$ una funzione di classe C^n in $[a, b]$. Allora esso è l'unico polinomio che soddisfa le condizioni $P^{(n)}(x_0) = f^{(n)}(x_0)$.

La dimostrazione della formula di Taylor si può trovare in un qualsiasi testo di Analisi 1. Un modo alternativo e costruttivo per giungere alla formula è quello che parte dalla derivata $n-$esima di una funzione per poi risalire mediante l'utilizzo iterativo del calcolo integrale.
Possiamo, inoltre, notare che per $n = 1$ abbiamo l'approssimazione lineare di f vista in nel paragrafo precedente, cioè

$$f(x) \cong f(x_0) + f'(x_0)(x - x_0).$$

A questo punto, consideriamo $f(x) = e^x$ e sviluppiamola in serie di Taylor centrata in 0 fino al terzo ordine ($n = 3$), cioè ne tronchiamo lo sviluppo a $n = 3$:

$$e^x \cong 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!}.$$

In analogia a quando visto nel paragrafo precedente, usiamo quest'approssimazione per calcolare determinati valori di e^x in un intorno di $x = 0$.

- Per $x = 0{,}1$, l'approssimazione ci dà come risultato $1 + 0{,}1 + \frac{0{,}01}{2} + \frac{0{,}001}{6} = 1{,}105166 \ldots$ e il valore vero di $e^{0{,}1}$ è $e^{0{,}1} = 1{,}10517 \ldots$; l'errore che si commette è dell'ordine di 10^{-5}, cioè inferiore a quello commesso considerando l'approssimazione lineare (che era $5 \cdot 10^{-3}$).
- Per $x = 0{,}2$, l'approssimazione è $1 + 0{,}2 + \frac{0{,}04}{2} + \frac{0{,}008}{6} = 1{,}221333 \ldots$ e $e^{0{,}2} = 1{,}22140 \ldots$; l'errore è circa 10^{-4}, anche qui inferiore a quanto visto nel caso lineare.
- Per $x = 0{,}5$, l'approssimazione ci dà $1 + 0{,}5 + \frac{0{,}25}{2} + \frac{0{,}125}{6} = 1{,}6458 \ldots$ e $e^{0{,}5} = 1{,}6487 \ldots$; l'errore è di $0{,}0029 \ldots$ anche qui inferiore a quanto visto nel caso lineare.
- Per $x = 1$, l'approssimazione ci dà $1 + 1 + \frac{1}{2} + \frac{1}{6} = 2{,}6666 \ldots$ ma $e^1 = e = 2{,}71828 \ldots$; l'errore è di circa $0{,}05$.
- Per $x = 5$, l'approssimazione ci dà $1 + 5 + \frac{25}{2} + \frac{125}{6} = 39{,}333 \ldots$ ma $e^5 = 148{,}413 \ldots$; analogamente al caso precedente, l'approssimazione dà un valore piuttosto distante da quello vero ed è inefficace.

Una breve descrizione dell'errore della formula di Taylor verrà trattata nei prossimi paragrafi. Tuttavia, riferendoci al caso dell'approssimazione lineare visto in precedenza, tenendo fede anche a quest'ultimo esempio, possiamo ricavare delle utili considerazioni.

- La formula di Taylor è efficace localmente, cioè approssima in maniera *utile* una qualsiasi funzione sufficientemente regolare (C^n) con un polinomio ma, come preannunciato, questo modo di operare si rivela essere sempre più inefficace via via che ci si allontana dal centro dello sviluppo.
- La formula di Taylor approssima la f sempre meglio quanti più termini dello sviluppo si considerano. La serie troncata al nono ordine è molto meglio di quella al terzo, ad esempio.

Ora sviluppiamo secondo la formula di Taylor alcune funzioni elementari. Se $f(x) = e^x$, risulta $f^{(n)}(x) = e^x$ per ogni n. Quindi, ponendo $x_0 = 0$, si ha $f^{(n)}(0) = e^0 = 1$ per ogni n. Perciò

$$e^x = 1 + x + \frac{x^2}{2} + \ldots + \frac{x^n}{n!} + R_n(x).$$

Analogamente, scegliendo $x_0 = 0$, si ottiene

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots + (-1)^{n+1}\frac{x^n}{n} + R_n(x);$$

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + R_{2n+1}(x);$$

$$\cos(x) = 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \ldots + (-1)^n \frac{x^{2n}}{(2n)!} + R_{2n}(x).$$

Nelle seguenti sezioni ci sarà una breve trattazione per quanto riguarda il resto nella formula di Taylor. La formula di Taylor consente di approssimare una funzione di classe C^n con un polinomio e i polinomi sono le funzioni più semplici in matematica: avere una stima del resto consente di approssimare con un errore arbitrario i valori di funzioni arbitrarie. Questo argomento – ovvero la tabulazione di funzioni – sarà trattato brevemente nell'ultima sezione di questo capitolo.

1.3.3 Resto di Peano (o piccolo)

Definiamo la funzione resto:

$$R_n(x) = f(x) - p_n(x).$$

La funzione $R_n(x)$ è il resto della formula di Taylor di f e rappresenta l'errore che si commette quando in x si sostituisce a $f(x)$ il suo polinomio di Taylor di centro x_0 e ordine n. Ora, se f è di classe C^n in un intorno di x_0, il resto $R_n(x)$ è un infinitesimo in x_0 di ordine superiore a $(x - x_0)^n$, ossia

$$\lim_{x \to x_0} \frac{R_n(x)}{(x - x_0)^n} = 0.$$

Siano $f(x), g(x)$ funzioni definite in un intorno di x_0 (con la eventuale eccezione di x_0), non nulle per $x \neq x_0$ e ambedue infinitesime per $x \to x_0$. Diremo che $f(x)$ è per $x \to x_0$ un infinitesimo di ordine superiore a $g(x)$, oppure equivalentemente che $f(x)$ è un "*o* piccolo" di $g(x)$, e cioè

$$f(x) = \mathrm{o}\big(g(x)\big), \qquad x \to x_0$$

se

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = 0.$$

Questa definizione introduce il resto secondo Peano ([14], §10) e si rappresenta anche nel modo seguente:

$$R_n(x) = \mathrm{o}((x - x_0)^n), \qquad x \to x_0.$$

Tuttavia, la notazione

$$f(x) = \mathrm{o}\big(g(x)\big)$$

è un abuso di scrittura e non è formalmente corretta: dobbiamo infatti mettere in guardia che l'uguaglianza che vediamo non corrisponde all'usuale relazione tra funzioni poiché $\mathrm{o}\big(g(x)\big)$ rappresenta una classe di funzioni e non una singola funzione. Sarebbe più corretto scrivere $f(x) \in \mathrm{o}\big(g(x)\big)$ ma anche qui ci serviremo, per comodità, dell'usuale notazione presentata in gran parte dei libri di testo.

Tenendo presenti le espressioni del resto e del polinomio di Taylor, utilizzando il simbolo di sommatoria, si può scrivere la formula di Taylor con il resto di Peano nella forma:

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + \mathrm{o}((x - x_0)^n).$$

Se, invece, consideriamo lo sviluppo di Taylor centrato nello zero, cioè la formula di McLaurin

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + \ldots + \frac{f^{(n)}(0)}{n!}x^n + \mathrm{o}(x^n).$$

La definizione del resto di Peano – cioè l'*o* piccolo – è piuttosto formale e, con questa scrittura, indichiamo delle proprietà teoriche piuttosto che un calcolo pratico. Essa, infatti, racchiude delle definizioni di limite e non dà una stima precisa del resto nella formula di Taylor.

Vediamo tuttavia alcune delle proprietà dell'*o* piccolo ([14], §10):

- $\mathrm{o}(x^n) + \mathrm{o}(x^n) = \mathrm{o}(x^n)$;
- $c \cdot \mathrm{o}(x^n) = \mathrm{o}(cx^n) = \mathrm{o}(x^n), \qquad c \in \mathbb{R}\backslash\{0\}$;
- $\mathrm{o}(x^n) - \mathrm{o}(x^n) = \mathrm{o}(x^n)$;
- $\mathrm{o}(x^m) \cdot \mathrm{o}(x^n) = \mathrm{o}(x^{m+n})$;
- $x^m \cdot \mathrm{o}(x^n) = \mathrm{o}(x^{m+n})$;
- $\mathrm{o}\big(\mathrm{o}(x^n)\big) = \mathrm{o}(x^n)$;
- $\mathrm{o}\big(x^n + \mathrm{o}(x^n)\big) = \mathrm{o}(x^n)$.

1.3.4 *O* grande

La notazione di O grande è formalmente simile a quella dell'o piccolo vista nel paragrafo precedente salvo indicare una relazione di uguaglianza di infiniti (o infinitesimi) invece dell'ordine superiore espresso dall'o piccolo.
Siano $f(x)$ e $g(x)$ due funzioni a valori reali e sia $x_0 \in \mathbb{R}$ una costante fissata. Diremo che $f(x)$ è un "O grande" di $g(x)$ se $\forall x > x_0$ risulta ([20])

$$|f(x)| \leq M \cdot |g(x)|, \qquad M \in \mathbb{R}$$

e si indica con $f(x) = \mathrm{O}\big(g(x)\big)$.
L' O grande nasce soprattutto per lo studio asintotico del comportamento delle funzioni all'infinito anche se, in seguito, venne adattato allo studio degli infinitesimi in analogia con l'o piccolo.
Diremo che $f(x)$ è per $x \to x_0$ un infinitesimo dello stesso ordine di $g(x)$, oppure equivalentemente che $f(x)$ è un "O grande" di $g(x)$, e cioè

$$f(x) = \mathrm{O}\big(g(x)\big), \qquad x \to x_0$$

se $g(x)$ è una funzione infinitesima per $x \to x_0$ e

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = l \in \mathbb{R}\backslash\{0\}.$$

La definizione appena data è coerente con la trattazione del resto nella formula di Taylor e anche in questo caso c'è un abuso di scrittura nell'utilizzo del simbolo di uguaglianza poiché $\mathrm{O}\big(g(x)\big)$ rappresenta una classe di funzioni e non una singola funzione.
A questo punto, volendo rappresentare la formula di Taylor con il resto servendoci della notazione di O grande in luogo dell'o piccolo otterremo

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + \mathrm{O}((x - x_0)^{n+1}).$$

Se, invece, consideriamo lo sviluppo di Taylor centrato nello zero, cioè la formula di McLaurin

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + \ldots + \frac{f^{(n)}(0)}{n!}x^n + O(x^{n+1}).$$

Così come l'o piccolo, anche la definizione di resto tramite l'O grande è piuttosto formale poiché indica delle proprietà teoriche piuttosto che un calcolo pratico. Inoltre, generalmente, la relazione di O grande è intesa nel senso "dello stesso ordine di..." sia in termini di infiniti che di infinitesimi.
Per l'O grande, inoltre, valgono analoghe proprietà dell'o piccolo:

- $O(x^n) + O(x^n)$;
- $c \cdot O(x^n) = O(xc^n) = O(x^n), \qquad c \in \mathbb{R}\backslash\{0\}$;
- $O(x^n) - O(x^n) = O(x^n)$;
- $O(x^m) \cdot O(x^n) = O(x^{m+n})$;
- $x^m \cdot O(x^n) = O(x^{m+n})$;
- $O\big(O(x^n)\big) = O(x^n)$;
- $O\big(x^n + O(x^n)\big) = O(x^n)$.

Come già detto, queste definizioni di resto sono piuttosto formali anche se risulteranno decisamente comode nella valutazione di funzioni anche complicate. Dire, ad esempio,

$$f(x) = O\big(g(x)\big),$$

per $x \to +\infty$, significa che, per quanto incomprensibile possa essere tale funzione, essa, all'infinito, si comporta come $g(x)$ (in genere più semplice di quella di partenza).

1.3.5 Resto integrale ([14], §10)

Se f è di classe C^1 in $[a, b]$, il resto $R_n(x)$ si può rappresentare nella forma

$$R_n(x) = \int_{x_0}^{x} \frac{(x-t)^n}{n!} f^{(n+1)}(t)dt\,, \qquad \forall x \in [a, b].$$

Esso viene chiamato resto integrale della formula di Taylor.

1.3.6 Resto di Lagrange ([14], §10)

Se f è di classe C^1 in $[a, b]$, $\forall x \in [a, b]$ esiste un numero $t \in [a, b]$, tale che

$$R_n(x) = \frac{f^{(n+1)}(t)}{(n+1)!}(x - x_0)^{n+1}.$$

Esso è il resto di Lagrange della formula di Taylor.

1.3.7 Tabulazione di funzioni ([14], §10)

Un utilizzo pratico della formula di Taylor è la tabulazione numerica di funzioni: si approssima un valore di una funzione $f(x)$ con un polinomio di Taylor di grado n, scegliendo x_0 ed n in modo tale che il resto $R_n(x)$ sia compatibile con il grado di precisione consentito dal problema.

A questo scopo è necessario avere una stima del resto R_n. Sia $f(x)$ una funzione di classe C^n in $[a, b]$ e $x_0 \in [a, b]$. Posto

$$M_{n+1} = \max\{|f^{(n+1)}(x)| : x \in [a, b]\},$$

il resto $R_n(x)$ della formula di Taylor verifica la disuguaglianza

$$|R_n(n)| \leq M_{n+1} \cdot \frac{|x - x_0|^{n+1}}{(n+1)!}, \qquad \forall x \in [a, b].$$

Diamo un'indicazione di come calcolare numericamente i valori di una data funzione usando la formula di Taylor. Questo metodo è utilizzato dagli elaboratori per calcolare i valori di funzioni non polinomiali e, grazie ad esso, si possono costruire tavole di logaritmi o di funzioni trigonometriche.

Innanzitutto si fissa il numero di cifre decimali con cui lavorare o, meglio, il grado di precisione con cui si vuole conoscere il risultato. Si esegue il calcolo usando il polinomio di

Taylor, trascurando il resto $R_n(x)$. Il resto, o errore che si commette, è stimato con la formula vista in precedenza.

Vediamo di fare un esempio: calcoliamo ora valori numerici approssimati del numero di Nepero; utilizziamo la formula di Taylor per la funzione $f(x) = e^x$, con centro $x_0 = 0$ e $x = 1$.

Poiché la derivata $f^{(n)}(x)$ è uguale ad e^x qualunque sia n, e dato che la funzione e^x è strettamente crescente, risulta

$$M_{n+1} = \max\{e^x : x \in [0,1]\} = e < 3.$$

Ponendo $x_0 = 0, x = 1$, nella stima del resto abbiamo

$$|R_n(x)| \leq M_{n+1} \frac{1}{(n+1)!} < \frac{3}{(n+1)!}.$$

Ad esempio, per $n = 10$ si trova $|R_n(x)| < \frac{3}{11!} < 10^{-7}$. Quindi otteniamo il numero e dal polinomio di Taylor per la funzione e^x con $x_0 = 0, x = 1, n = 10$:

$$e \cong 1 + 1 + \frac{1}{2} + \frac{1}{3!} + \frac{1}{4!} + \ldots + \frac{1}{10!} = 2.71828180114 \ \ldots$$

Il risultato è stato ottenuto a meno di un errore inferiore a 10^{-7}. Perciò il valore di e esatto fino alla sesta cifra decimale è il seguente:

$$e = 2{,}718281 \ \ldots$$

Per valori più grandi di n con lo stesso metodo si trovano, ad es, le prime venti cifre decimali:

$$e = 2{,}71828182845904523536 \ \ldots$$

2. RICHIAMI DI ANALISI MATEMATICA II

In questa sezione si rivedranno concetti di Analisi Matematica II propedeutici alla comprensione, soprattutto, della successiva sezione di Analisi Complessa. Vedremo, infatti, che molti risultati di quest'ultima sono direttamente riconducibili a teoremi di Analisi II.
Si inizierà a considerare brevemente successioni e serie di funzioni che serviranno per ampliare un po' gli orizzonti inquadrati con successioni e serie numeriche. Poi, dopo una breve introduzione nella quale si vedranno concetti basilari di Topologia, si passerà ad una trattazione delle funzioni a due variabili e delle curve del piano.

2.1 SUCCESSIONI DI FUNZIONI

2.1.1 Definizioni preliminari

In questo paragrafo tratteremo le successioni di funzioni.
Esse si differenziano da quelle viste in precedenza per il fatto che, invece che con numeri reali (o complessi), abbiamo a che fare con delle funzioni. Questa differenza si tradurrà nei risultati e in uno studio più approfondito che dovrà considerare una convergenza non più semplice ma, in un certo senso, condizionata. Infatti, si parlerà di "intervallo" di convergenza poiché dovremo tenere conto dell'insieme dei valori assunti dalla variabile per i quali è possibile avere una convergenza.
Diremo che una successione – o sequenza – di funzioni, è una particolare applicazione che ad ogni numero naturale associa una funzione.

$$n \in \mathbb{N} \longmapsto f_n(x)$$

In analogia alle successioni numeriche, può accadere che l'indice n debba talora restringere il suo ambito agli interi positivi oppure ai naturali maggiori di un fissato valore N.
Per esempio la successione di funzioni

$$f_n(x) = x^n, \qquad (n \in \mathbb{N})$$

al variare di $n \in \mathbb{N}$ associa ad n la funzione di variabile reale $x \longmapsto x^n$: quindi i suoi valori saranno, di volta in volta, le funzioni $f_0(x) = 1, f_1(x) = x, f_2(x) = x^2, \ldots, f_n(x) = x^n, \ldots$.
Nella successione appena vista non ci sono quindi problemi con gli indici. Se, invece, avessimo avuto a che fare con una successione di funzioni del tipo

$$f_n(x) = \frac{x}{n}, \qquad (n \in \mathbb{N}\backslash\{0\})$$

prenderemmo atto che essa è definita per tutti i naturali escluso il caso $n = 0$.

Possiamo già notare che il discorso è più complicato rispetto alle successioni numeriche. Occorre, infatti, vedere il comportamento di $f_n(x)$ che varia al variare di $x \in \mathbb{R}$ oltre che $n \in \mathbb{N}$. Questi argomenti verranno affrontati nel paragrafo successivo.
Per indicare le successioni di funzioni, ci si serve di notazioni simili alle successioni numeriche: avremo dunque scritture come $\left(f_n(x)\right)_{n\in\mathbb{N}}$ o anche $(f_n)_{n\in\mathbb{N}}$ come abbreviazione.
Sia ora $I \subseteq \mathbb{R}$ e $f_n: I \to \mathbb{R}$ una successione di funzioni reali definite in I. Diremo che $(f_n)_n$ converge puntualmente ([15] §1) in I alla funzione $f: I \to \mathbb{R}$ se $\forall x \in I$ risulta

$$\lim_{n\to+\infty} f_n(x) = f(x).$$

Tenendo fede alla definizione di limite, $f_n(x) \to f(x)$ se, per ogni $\varepsilon > 0$ e $x \in I$, esiste $n_0 \in \mathbb{N}$ t.c.

$$f(x) - \varepsilon < f_n(x) < f(x) + \varepsilon \qquad \forall n \in \mathbb{N}, \qquad n > n_0$$

che si può tradurre anche con

$$|f_n(x) - f(x)| < \varepsilon \qquad \forall n \in \mathbb{N}, \qquad n > n_0.$$

Ora occorre fare una precisazione. Dire che $(f_n)_n$ converge puntualmente, equivale ad affermare che per ogni x fissato, avviene la convergenza: l'aggettivo "puntualmente" è inteso proprio nel senso di "punto per punto" (dall'inglese *pointwise*). Inoltre, generalmente, n_0 non è assoluto, ma dipende da x: a questo proposito possiamo, per esempio, considerare la successione di funzioni

$$f_n(x) = x + \frac{x}{n}, \qquad n \geq 1.$$

Essa converge puntualmente a $f(x) = x$. Fissiamo $\varepsilon = 0{,}5$.

- Se $x = 1$, $n_0 = 2$ poiché $|f_n(1) - f(1)| = |1 + 1/n - 1| = |1/n| < 0{,}5$ per $n > 2$.
- Se $x = 10$, $n_0 = 20$ poiché $|f_n(10) - f(10)| = |10 + 10/n - 10| = |10/n| < 0{,}5$ per $n > 20$.
- Se $x = 0{,}1$, $n_0 = 0$ poiché $|f_n(0{,}1) - f(0{,}1)| = |0{,}1 + 0{,}1/n - 0{,}1| = |0{,}1/n| < 0{,}5$ per $n \geq 1$.

L'esempio appena visto serve proprio a chiarire la differenza che intercorre tra sequenze numeriche e successioni di funzioni. Si può notare che, fissato un valore di $x_0 \in I$, la successione di funzioni si riduce ad una successione numerica $\left(f_n(x_0)\right)_{n\in\mathbb{N}}$.
La complicazione principale è proprio il fatto che c'è la variabile $x \in \mathbb{R}$ da considerare per stabilire il carattere della successione di funzioni per cui in genere la convergenza non è una unica: ci può essere infatti convergenza per alcuni valori di x e per altri no.
Nell'esempio precedente si è visto come, generalmente, n_0 non sia unico ma dipenda da x. Diremo allora che $f_n(x)$ converge uniformemente a f in I se $\forall \varepsilon > 0$, $\exists n_0 \in \mathbb{N}$ dipendente solo da ε tale che

$$|f_n(x) - f(x)| < \varepsilon \qquad \forall n \in \mathbb{N}, \qquad n > n_0, \qquad \forall x \in I.$$

La convergenza uniforme è una condizione più forte di quella puntuale: possiamo dire che in questo caso – oltre alla semplice convergenza per $x \in I$ – si richiede che il *modo* di convergere della funzione sia lo stesso in qualsiasi $x \in I$. Come tale, abbiamo dunque che la convergenza uniforme implica quella puntuale.

2.1.2 Risultati sulle successioni di funzioni

In questo paragrafo richiameremo alcuni risultati importanti sulle successioni di funzioni.

Teorema ([15], §1; [2], §11.3)
Sia $f_n(x): I \subseteq \mathbb{R} \to \mathbb{R}$ una successione di funzioni continue che converge uniformemente a f in I. Allora f è continua.

Ammesso che $f_n(x)$ sia una successione di funzioni continue se la convergenza è solo puntuale non è detto che f sia continua. Si può vedere che, ad esempio, la successione di funzioni continue

$$f_n(x) = x^n, \qquad x \in [0,1]$$

converge puntualmente alla funzione

$$f(x) = \begin{cases} 1, & x = 1 \\ 0, & 0 \leq x < 1 \end{cases}$$

che non è affatto continua in $[0,1]$.

Teorema (passaggio al limite sotto il segno di integrale) ([15], §1; [2], §11.4)
Sia $f_n(x)$ una successione di funzioni continue convergente uniformemente ad f in $[a, b]$, allora

$$\lim_{n\to\infty} \int_a^b f_n(x)dx = \int_a^b f(x)dx.$$

Teorema (passaggio al limite sotto il segno di derivata) ([15], §1)
Sia $f_n(x)$ una successione di funzioni di classe C^1 in $[a, b]$. Supponiamo che esista $x_0 \in [a, b]$ tale che la successione numerica $f_n(x_0)$ converga e che la successione delle derivate $f_n^{'}(x)$ converga uniformemente in $[a, b]$. Allora $f_n(x)$ converge uniformemente a f di classe C^1 e risulta

$$\lim_{n\to\infty} f_n{}'(x) = f^{'}(x).$$

2.2 SERIE DI FUNZIONI

2.2.1 Introduzione

Si introducono le serie di funzioni allo stesso modo in cui partendo dalle successioni numeriche si arriva a quelle di funzioni. Definiremo, dunque, la serie di funzioni come serie infinita (salvo eventuali precisazioni *in loco*)

$$\sum_{k=0}^{\infty} f_k(x).$$

Quest'ultima è detta somma della serie di termine generale f_k ([15], §1). Quindi non è altro che una serie i cui elementi sono funzioni, anziché numeri. Esisterà allora un insieme di definizione delle f_k (spesso un intervallo nei casi che ci interesseranno) che chiameremo $I \subseteq \mathbb{R}$ (o $[a, b]$ o (a, b), con $a, b \in \mathbb{R}$). In analogia con le serie numeriche, possiamo definire la successione delle somme parziali

$$S_n(x) = \sum_{k=0}^{n} f_k(x).$$

Si dice che la serie di funzioni converge puntualmente a f in I se, per ogni $x \in I$,

$$f(x) = \lim_{n\to\infty} S_n(x) = \lim_{n\to\infty} \sum_{k=0}^{n} f_k(x) = \sum_{k=0}^{\infty} f_k(x).$$

Diremo, inoltre, che essa è assolutamente convergente in I se la serie dei valori assoluti – cioè quella di termine generale $|f_k(x)|$ – converge puntualmente su I.
Inoltre, se esiste una successione di numeri reali non negativi a_k tali che $|f_k(x)| \leq a_k$ per ogni $x \in I$, $k \in \mathbb{N}$ e $\sum a_k$ converge allora diremo che la serie $f_k(x)$ converge totalmente in I.

2.2.2 Risultati sulle serie di funzioni

In analogia con le serie numeriche, anche per le serie di funzioni continuano a valere gli stessi criteri di convergenza (confronto, radici, …). In questo paragrafo saranno trattati dei risultati utili per ciò che concerne le serie di funzioni.

Teorema
La convergenza totale di una serie di funzioni implica la convergenza uniforme.

Teorema (continuità della somma) ([15], §1)
Sia $\sum_{k=0}^{\infty} f_k(x)$ una serie di funzioni continue. Se $\sum_{k=0}^{\infty} f_k(x) \to f(x)$ uniformemente in I allora la funzione f è continua.

Teorema (integrazione per serie)
Sia $\sum_{k=0}^{\infty} f_k(x)$ una serie di funzioni continue in $[a, b]$. Se essa converge uniformemente ad una f continua in $[a, b]$, allora

$$\int_a^b f(x)dx = \int_a^b \sum_{k=0}^{\infty} f_k(x)\, dx = \sum_{k=0}^{\infty} \int_a^b f_k(x)dx.$$

Teorema (derivazione per serie) ([15], §1)
Sia $f_k(x)$ una successione di funzioni di classe C^1 in $[a, b]$. Se $\sum_{k=0}^{\infty} f_k(x)$ converge uniformemente a f allora f è anch'essa di classe C^1 in $[a, b]$ e risulta

$$f'(x) = \sum_{k=0}^{\infty} f_k'(x), \qquad \forall x \in [a,b]$$

ammesso che anche la serie delle derivate converga uniformemente in $[a,b]$.

2.3 SERIE DI POTENZE

2.3.1 Introduzione

Le serie di potenze sono particolari tipi di serie di funzione. La loro importanza si ritrova soprattutto per le funzioni di variabile complessa che saranno trattate nella successiva sezione di richiami. La peculiarità delle serie di potenze, invece, sta nei risultati riguardanti la convergenza delle stesse che sono più semplici rispetto a quelli di generiche serie di funzioni.
Inizialmente, consideriamo la serie geometrica, già accennata nella sezione di richiami di analisi I.

$$\sum_{k=0}^{\infty} x^k = 1 + x + x^2 + x^3 + \cdots + x^n + \cdots$$

Nella sezione precedente, l'avevamo vista come particolare serie numerica per x fissato mentre ora la consideriamo come una serie di funzioni che converge per $|x| < 1$, cioè $x \in (-1,1)$.
Più in generale, una serie di potenze è definita nel seguente modo

$$\sum_{k=0}^{\infty} a_k\,(x - x_0)^k$$

nel quale a_k è una successione numerica mentre x_0 è un numero reale fissato. La chiameremo serie di potenze centrata nel punto x_0. Possiamo concludere che la serie geometrica non è altro che un particolare tipo di serie di potenze nella quale $a_k = 1$ per ogni k mentre $x_0 = 0$.
L'insieme dei valori per i quali una tale serie converge è sempre del tipo $(x_0 - r, x_0 + r)$, che per questo motivo viene chiamato intervallo di convergenza mentre r è detto raggio di convergenza ([2], §11.6). Possiamo notare che una qualsiasi serie di potenze converge sempre almeno in un punto, cioè per $x = x_0$, infatti

$$\sum_{k=0}^{\infty} a_k (x_0 - x_0)^k = \sum_{k=0}^{\infty} a_k \cdot 0^k = a_0, \qquad \forall a_k$$

(In essa si pone $0^0 = 1$).
Un esempio importante di serie di potenze è quello coinvolto nella formula di Taylor, che è la generalizzazione della formula vista nella precedente sezione di richiami (§1.3.2)

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k, \qquad con\ f \in C^{\infty};$$

essa è chiamata anche serie di Taylor e, in essa, $a_k = f^{(k)}(x_0)/k! \in \mathbb{R}$. La differenza con quanto visto nella sezione precedente è che qui la formula di Taylor non è troncata e, invece del resto (Peano, Lagrange o integrale), troviamo la serie nella sua forma completa.
Non è riduttivo lavorare con serie di potenze della forma

$$\sum_{k=0}^{\infty} a_k x^k,$$

ovvero assumere $x_0 = 0$: basta operare un semplice cambio di variabile $y = x - x_0$. Solo per semplicità, la nuova variabile la chiameremo x invece di y. Tutti i risultati del prossimo paragrafo saranno riferiti a serie di questo tipo: se si ha a che fare con una serie di potenze generica basta operare il cambio di variabile appena suggerito ($y = x - x_0$) e – ottenuto ciò che si cercava – tornare indietro con il cambio di variabile inverso ($x = y + x_0$).

2.3.2 Risultati sulle serie di potenze

Elenchiamo in questo paragrafo alcuni risultati importanti sulle serie di potenze.

Teorema
Ogni serie di potenze $\sum_{k=0}^{\infty} a_k x^k$ soddisfa sempre una (e una sola) delle seguenti condizioni:

i) la serie converge solo per $x = 0$;
ii) la serie converge $\forall x \in \mathbb{R}$;
iii) $\exists r \in \mathbb{R}$, $0 < r < +\infty$ tale che la serie converge per $|x| < r$ e diverge per $|x| > r$.

Teorema
Se la serie di potenze $\sum_{k=0}^{\infty} a_k x^k$ converge per $x = a$, con a reale non nullo, allora essa converge totalmente in ogni intervallo chiuso e limitato incluso nell'intervallo

$$(-|a|, |a|)$$

e quindi anche assolutamente e uniformemente in ogni tale intervallo.

Questo risultato è molto interessante: la convergenza uniforme in $(-|a|, |a|)$ implica che, detta f la funzione cui si converge, essa sarà continua in tale intervallo per il teorema di continuità del limite visto per serie di funzioni ([2], §11.6).

Teorema
Consideriamo due serie di potenze $\sum_{k=0}^{\infty} a_k x^k$ e $\sum_{k=0}^{\infty} b_k x^k$. Se, nel loro intervallo di convergenza, hanno come somma la stessa funzione f, allora sono uguali termine a termine e risulta

$$a_k = b_k = \frac{f^{(k)}(0)}{k!} \qquad \forall k \geq 0.$$

Teorema (D'Alembert)
Sia data la serie di potenze $\sum_{k=0}^{\infty} a_k x^k$ con $a_k \neq 0$ per ogni $k \in \mathbb{N}$. Se esiste il limite

$$l = \lim_{k\to\infty} \left|\frac{a_{k+1}}{a_k}\right|$$

allora il raggio di convergenza della serie è uguale a

$$r = \begin{cases} +\infty & \text{se } l = 0, \\ \dfrac{1}{l} & \text{se } 0 < l < +\infty, \\ 0 & \text{se } l = +\infty, \end{cases}$$

Teorema (Cauchy-Hadamard)

Sia data la serie di potenze $\sum_{k=0}^{\infty} a_k x^k$. Se esiste il limite

$$l = \lim_{k\to\infty} \sqrt[k]{|a_k|},$$

allora il raggio di convergenza della serie è

$$r = \begin{cases} +\infty & \text{se } l = 0, \\ \dfrac{1}{l} & \text{se } 0 < l < +\infty, \\ 0 & \text{se } l = +\infty. \end{cases}$$

Teorema

Indichiamo con $\sum_{k=1}^{\infty} k a_k x^{k-1}$ la serie derivata della serie di potenze $\sum_{k=0}^{\infty} a_k x^k$, cioè quella che si ottiene derivando termine a termine i suoi addendi con le usuali regole di derivazione. Allora la serie di potenze ha lo stesso raggio di convergenza della serie derivata.

Teorema (derivazione e integrazione per serie di potenze)

Se la serie di potenze $\sum_{k=0}^{\infty} a_k x^k$ ha un raggio di convergenza r non nullo e se $f(x)$ è la sua somma, cioè

$$f(x) = \sum_{k=0}^{\infty} a_k x^k \qquad \forall |x| < r,$$

allora f è derivabile e risulta

$$f'(x) = \sum_{k=1}^{\infty} k a_k x^{k-1}, \qquad \forall |x| < r.$$

Inoltre f è anche integrabile e risulta

$$\int_0^x f(t)dt = \sum_{k=0}^{\infty} \frac{a_k}{k+1} x^{k+1}, \qquad \forall |x| < r.$$

Teorema

Se la serie di potenze

$$\sum_{k=0}^{\infty} a_k (x - x_0)^k$$

ha raggio di convergenza $r > 0$, la sua somma $f(x)$ è una funzione di classe C^∞.

Inoltre f è sviluppabile in serie di Taylor nel modo usuale con $a_k = f^{(k)}(x_0)/k!$.

2.4 CENNI DI GEOMETRIA ANALITICA E TOPOLOGIA

Nei prossimi paragrafi saranno esaminati concetti importanti circa le funzioni a due variabili che si riveleranno utili, soprattutto, per la successiva sezione di richiami di analisi complessa. Tuttavia si tratta di oggetti qualitativamente differenti dalle usuali funzioni a una variabile con cui abbiamo avuto a che fare fino ad ora e la loro analisi si basa su nozioni basilari di topologia e geometria analitica.
Basandosi sull'usuale riferimento cartesiano – utilizzato in matematica fin dalle scuole superiori – si vedrà la caratterizzazione di $\mathbb{R}^2$ come spazio vettoriale per poi esaminare concetti elementari di topologia applicati allo spazio $\mathbb{R}^2$.

2.4.1 Introduzione

Con il simbolo $\mathbb{R}^2$, indichiamo l'insieme costituito dalle coppie ordinate di numeri reali [1]

$$\mathbb{R}^2 = \{(x_0, y_0): x_0, y_0 \in \mathbb{R}\}$$

Ogni elemento di $\mathbb{R}^2$ - detto comunemente *punto* – sarà proprio una coppia ordinata di numeri reali: possiamo rappresentarlo nell'usuale sistema di riferimento cartesiano ortogonale $0xy$ (Figura 2.1a).

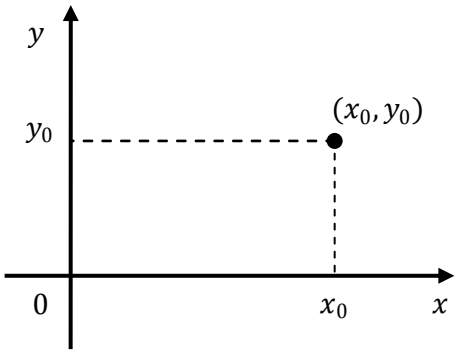

Figura 2.1a. Punti nel sistema di riferimento cartesiano ortogonale.

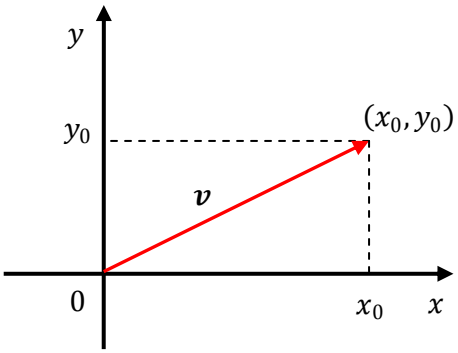

Figura 2.1b. Il vettore (x_0, y_0).

E' naturale identificare il punto (x_0, y_0) – cioè il punto di coordinate x_0, y_0 – col vettore $\boldsymbol{v}$ applicato all'origine degli assi, avente componenti x_0, y_0 (Figura 2.1b).
Definiamo la somma di due vettori $\boldsymbol{v_1} = (x_1, y_1)$ e $\boldsymbol{v_2} = (x_2, y_2)$ nel modo seguente: il vettore somma $\boldsymbol{v_1} + \boldsymbol{v_2}$ ha coordinate $x_1 + x_2, y_1 + y_2$, cioè

$$\boldsymbol{v_1} + \boldsymbol{v_2} = (x_1 + x_2, y_1 + y_2)$$

Inoltre, se α è una costante reale – detta anche uno scalare – possiamo definire il prodotto di un vettore $v = (x, y)$ per lo scalare α ponendo

$$\alpha \boldsymbol{v} = (\alpha x, \alpha y),$$

che è anch'esso un vettore.
Senza entrare in particolari necessari per una definizione tecnica appropriata, possiamo dire che $\mathbb{R}^2$ è uno spazio vettoriale su $\mathbb{R}$ ([23]). Infatti, l'insieme dei vettori $\mathbb{R}^2$ con le operazioni di somma e prodotto per uno scalare soddisfa gli assiomi della definizione di spazio vettoriale (sul campo reale).

2.4.2 Norma e prodotto scalare

Dato un vettore di $\mathbb{R}^2$, possiamo associare ad esso un numero reale detto norma o modulo del vettore. Il concetto di norma non è così semplice e servirebbe una trattazione più teorica; per gli obiettivi di questa tesi, però, ci atteniamo alla definizione di ([15], §2). La norma appena descritta è detta "norma euclidea" o "norma 2" ed è quella più utilizzata: infatti ci sono vari tipi di norme poiché con il termine norma si intende, in realtà, una funzione con determinate proprietà che associa numeri reali a dei vettori. In questo ambito, ci limitiamo comunque a definire la norma euclidea $\|.\|$ in $\mathbb{R}^2$ e fornirne anche alcune proprietà senza esaminare, nello specifico, il concetto generale di norma. Consideriamo allora

$$\|.\|: \mathbb{R}^2 \to \mathbb{R} \qquad \Longrightarrow \qquad \|\boldsymbol{v}\| = \sqrt{x^2 + y^2}, \qquad \forall \boldsymbol{v} = (x, y) \in \mathbb{R}^2.$$

Per la norma valgono le seguenti proprietà;

- $\|\boldsymbol{v}\| \geq 0, \forall \boldsymbol{v} \in \mathbb{R}^2$, inoltre $\|\boldsymbol{v}\| = 0 \Leftrightarrow \boldsymbol{v} = (0,0)$;
- $\|\boldsymbol{v}\| = \|-\boldsymbol{v}\|, \forall \boldsymbol{v} \in \mathbb{R}^2$;
- $\|\boldsymbol{v} + \boldsymbol{u}\| = \|\boldsymbol{u} + \boldsymbol{v}\|, \forall \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^2$;
- $\|\alpha \boldsymbol{v}\| = |\alpha| \|\boldsymbol{v}\|, \forall \alpha \in \mathbb{R}, \boldsymbol{v} \in \mathbb{R}^2$;
- $\|\boldsymbol{v} \pm \boldsymbol{u}\| \leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|, \forall \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^2$ (la disuguaglianza triangolare della norma).

Nell'usuale rappresentazione dei punti – o vettori – di $\mathbb{R}^2$ nel piano cartesiano, la norma (o modulo) del vettore $\boldsymbol{v} = (x_0, y_0)$ è proprio la distanza del punto (x_0, y_0) dall'origine degli assi $(0,0)$. Esso è detto anche lunghezza o intensità del vettore, come rappresentato in Figura 2.2.
Ricordiamo, a tal proposito, la formula per il calcolo della distanza (euclidea) tra due punti derivante proprio dal teorema di Pitagora. Con $P = (x_1, y_1)$ e $P' = (x_2, y_2)$, la distanza tra P e P' è

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

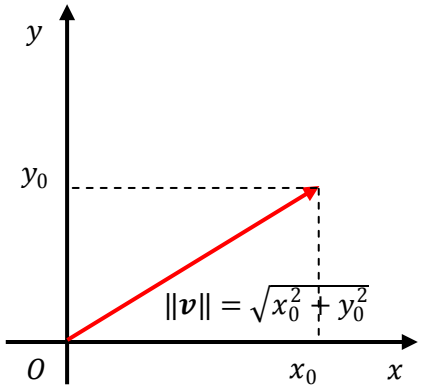

Figura 2.2. Rappresentazione geometrica della norma di $\boldsymbol{v}$.

Ponendo in essa $\boldsymbol{v} = (x_0, y_0)$ e come altro punto l'origine $O = (0,0)$ si ottiene proprio la formula di partenza: $\|\boldsymbol{v}\| = \sqrt{x_0^2 + y_0^2}$.
E' utile definire anche il prodotto scalare, strettamente legato alla norma appena esaminata. Dati due vettori $\boldsymbol{v_1} = (x_1, y_1)$, $\boldsymbol{v_2} = (x_2, y_2)$, indichiamo con $< \boldsymbol{v_1}, \boldsymbol{v_2} >$ il loro prodotto scalare, definito come

$$< \boldsymbol{v_1}, \boldsymbol{v_2} >= x_1 x_2 + y_1 y_2.$$

Il prodotto scalare, dunque, è una funzione che, applicata a coppie di vettori, restituisce un reale, cioè la somma dei prodotti delle corrispettive componenti. Nel caso in cui sia applicato allo stesso vettore (considerato 2 volte) otteniamo l'espressione che segue (e che lo collega alla norma)

$$< \boldsymbol{v}, \boldsymbol{v} >= x_0^2 + y_0^2 = \|\boldsymbol{v}\|^2, \qquad \forall \boldsymbol{v} = (x_0 y_0) \in \mathbb{R}^2.$$

Oltre alla proprietà appena vista, il prodotto scalare gode di altre proprietà non meno importanti.

- $< \boldsymbol{v}, \boldsymbol{u} >=< \boldsymbol{u}, \boldsymbol{v} >, \forall \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^2$
- $< \alpha \boldsymbol{u}, \boldsymbol{v} >= \alpha \cdot < \boldsymbol{u}, \boldsymbol{v} >, \forall \alpha \in \mathbb{R}, \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^2$
- $< \boldsymbol{u} + \boldsymbol{u}', \boldsymbol{v} >=< \boldsymbol{u}, \boldsymbol{v} > +< \boldsymbol{u}', \boldsymbol{v} >, \forall \boldsymbol{u}, \boldsymbol{u}', \boldsymbol{v} \in \mathbb{R}^2.$

Come esempio dei concetti di norma e prodotto scalare appena enunciati, supponiamo di avere $\boldsymbol{u} = (1, -2), \boldsymbol{u}' = (-2,3)$ e $\boldsymbol{v} = (2,0)$ tre vettori in $\mathbb{R}^2$.

- $\|\boldsymbol{u}\| = \sqrt{1^2 + (-2)^2} = \sqrt{1+4} = \sqrt{5} \geq 0,$
- $\|\boldsymbol{v}\| = \sqrt{2^2 + 0^2} = \sqrt{4} = 2 \geq 0,$
- $< \boldsymbol{u}, \boldsymbol{v} >= 1 \cdot 2 + (-2) \cdot 0 = 2,$
- $< \boldsymbol{u}, \boldsymbol{u} >= 1 \cdot 1 + (-2) \cdot (-2) = 1^2 + (-2)^2 = \|\boldsymbol{v}\|^2,$
- $\|\boldsymbol{u} + \boldsymbol{v}\| = \sqrt{(1+2)^2 + (-2+0)^2} = \sqrt{3^2 + (-2)^2} = \sqrt{13} \leq 2 + \sqrt{5} = \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$, inoltre $\|\boldsymbol{u} + \boldsymbol{v}\| = \sqrt{(1+2)^2 + (-2+0)^2} = \sqrt{(2+1)^2 + (0-2)^2} = \|\boldsymbol{v} + \boldsymbol{u}\|,$
- $\|\alpha \boldsymbol{u}\| = \sqrt{(\alpha \cdot 1)^2 + \big(\alpha \cdot (-2)\big)^2} = \sqrt{\alpha^2 \cdot 1^2 + \alpha^2 \cdot (-2)^2} = \sqrt{\alpha^2 \cdot (1^2 + (-2)^2)} = |\alpha| \cdot \sqrt{1^2 + (-2)^2} = |\alpha| \cdot \|\boldsymbol{u}\|, \forall \alpha \in \mathbb{R}^2,$
- $< \alpha \boldsymbol{u}, \boldsymbol{v} \geq \alpha \cdot (1 \cdot 2) + \alpha \cdot \big((-2) \cdot 0\big) = \alpha \cdot (1 \cdot 2 + (-2) \cdot 0) = \alpha \cdot < \boldsymbol{u}, \boldsymbol{v} >,$
- $< u + \boldsymbol{u}', \boldsymbol{v} >= (1-2) \cdot 2 + (-2+3) \cdot 0 = 1 \cdot 2 + (-2) \cdot 2 + (-2) \cdot 0 + 3 \cdot 0 = 1 \cdot 2 + (-2) \cdot 0 + (-2) \cdot 2 + 3 \cdot 0 =< \boldsymbol{u}, \boldsymbol{v} > +< \boldsymbol{u}', \boldsymbol{v} >.$

2.4.3 Un po' di topologia in $\mathbb{R}^2$

Richiamiamo adesso alcuni concetti topologici riferiti allo spazio $\mathbb{R}^2$.

Consideriamo $(x_0, y_0) \in \mathbb{R}^2$ e r reale positivo. Chiameremo cerchio aperto (o intorno circolare) di centro (x_0, y_0) e raggio r l'insieme

$$I_r = \left\{(x, y) \in \mathbb{R}^2 : \sqrt{(x - x_0)^2 + (y - y_0)^2} < r\right\},$$

mentre con la scrittura

$$C_r = \{(x, y) \in \mathbb{R}^2 : (x - x_0)^2 + (y - y_0)^2 = r^2\},$$

indicheremo la circonferenza di (x_0, y_0) e raggio r (il bordo di I_r).

Adottando la notazione di tipo vettoriale e indicando $\boldsymbol{v_0} = (x_0, y_0)$ e $\boldsymbol{v} = (x, y)$, abbiamo

$$I_r = \{\boldsymbol{v} \in \mathbb{R}^2 : \|\boldsymbol{v} - \boldsymbol{v_0}\| < r\}.$$

Nella sezione di richiami di analisi complessa, sarà anche utilizzato il termine "disco" piuttosto che "intorno circolare", tuttavia il significato è lo stesso anche se si adopererà una differente notazione

$$D(\boldsymbol{v_0}, r) = \{\boldsymbol{v} \in \mathbb{R}^2 : \|\boldsymbol{v} - \boldsymbol{v_0}\| < r\}$$

Per denotare il disco aperto di centro $\boldsymbol{v_0}$ e raggio r.

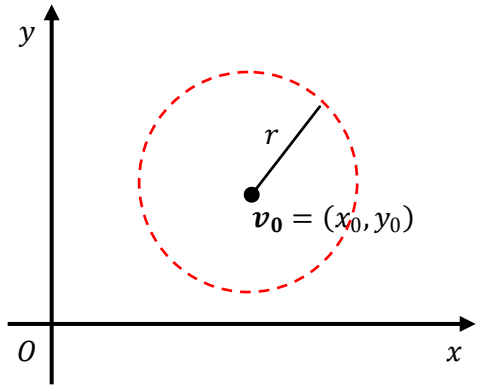

Figura 2.3. Intorno circolare o disco aperto di centro $\boldsymbol{v_0}$ e raggio r.

Facciamo ora riferimento alle Figure 2.4a, 2.4b e 2.4c. Consideriamo un punto $(x_0, y_0) \in \mathbb{R}^2$ e un sottoinsieme A di $\mathbb{R}^2$.

- Diremo che (x_0, y_0) è interno ad A se esiste un disco aperto di centro (x_0, y_0) contenuto in A.
- Diremo che (x_0, y_0) è un punto esterno ad A se esiste un disco aperto di centro (x_0, y_0) contenuto nel complementare di A (quindi (x_0, y_0) è interno al complementare di A).
- Diremo che (x_0, y_0) è un punto di frontiera per A se in ogni disco aperto di centro (x_0, y_0) si trovano punti interni sia di A che del complementare di A. In quel caso (x_0, y_0) non è né interno né esterno e – in termini non molto tecnici (si faccia riferimento alla Figura 2.4c) – diremo anche che (x_0, y_0) giace sul bordo di A.

L'insieme dei punti di frontiera forma la frontiera – o bordo – di A e la si indica con ∂A.

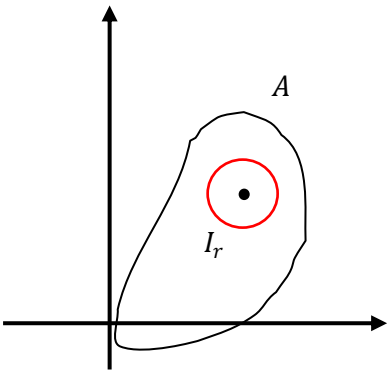

Figura 2.4a. Punto interno.

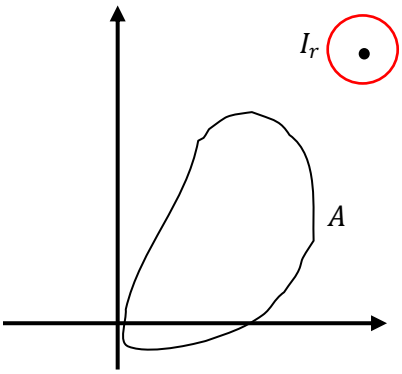

Figura 2.4b. Punto esterno.

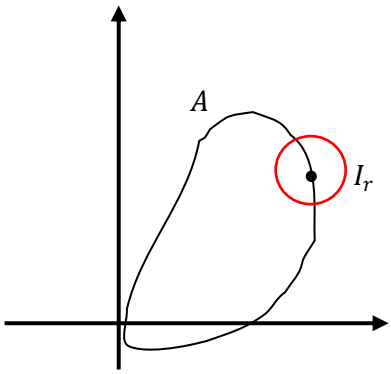

Figura 2.4c. Punto di frontiera.

Un sottoinsieme $A \subseteq \mathbb{R}^2$ si dice aperto se $\forall (x_0, y_0) \in A$ esiste un disco aperto di centro (x_0, y_0) contenuto in A o, in altre parole, ogni punto di A è anche un punto interno di A. Viceversa A si dice chiuso se il suo complementare (rispetto a $\mathbb{R}^2$) è aperto.

Diremo che (x_0, y_0) è un punto di accumulazione per l'insieme $A \subseteq \mathbb{R}^2$ se in ogni intorno circolare di (x_0, y_0) esiste almeno un punto di A diverso da (x_0, y_0); viceversa se ciò non accade allora il punto (x_0, y_0) si chiama un punto isolato per A.

Si può notare che ogni punto interno ad A è un punto di accumulazione.

La chiusura di un insieme A - che indicheremo con $\bar{A}$ – è l'insieme risultante dall'unione di A con i suoi punti di accumulazione. La chiusura $\bar{A}$ di A è un insieme chiuso e si può provare

che è anche l'unione tra A e l'insieme dei suoi punti di frontiera (il "bordo"). In particolare se A è chiuso, $\bar{A} = A$ poiché contiene già la sua frontiera mentre se A è aperto, allora

$$\bar{A} = A \cup \partial A.$$

Un dominio in $\mathbb{R}^2$ è la chiusura di un insieme aperto.
A questo punto possiamo definire il disco chiuso o l'intorno circolare chiuso di $\boldsymbol{v} = (x_0, y_0)$

$$\overline{D(\boldsymbol{v}, r)} = \bar{I}_r = \overline{\{\boldsymbol{v} \in \mathbb{R}^2 : \|\boldsymbol{v} - \boldsymbol{v_0}\| < r\}} = \{\boldsymbol{v} \in \mathbb{R}^2 : \|\boldsymbol{v} - \boldsymbol{v_0}\| \leq r\}$$

esso, dunque, è la chiusura del disco aperto già visto in precedenza ed è un dominio. L'ultima scrittura è giustificata dal fatto che consideriamo l'insieme dei punti distanti al massimo r dal centro e quindi prendiamo anche la circonferenza che delimita l'intorno circolare.
In forma vettoriale, la circonferenza di centro $\boldsymbol{v}$ e raggio r – che indicheremo con $C(\boldsymbol{v}, r)$ – è la seguente

$$C(\boldsymbol{v}, r) = \{\boldsymbol{v} \in \mathbb{R}^2 : \|\boldsymbol{v} - \boldsymbol{v_0}\| = r\} = \{(x, y) \in \mathbb{R}^2 : (x - x_0)^2 + (y - y_0)^2 = r^2\}.$$

Essa è il "bordo" del disco aperto di centro $\boldsymbol{v_0}$ e raggio r e rappresenta il luogo geometrico dei punti distanti r dal centro.

2.4.4 Proprietà e caratteristiche dei sottoinsiemi di $\mathbb{R}^2$

In precedenza abbiamo introdotto alcune definizioni preliminari per prendere dimestichezza con oggetti topologici di $\mathbb{R}^2$ quali aperti e chiusi. L'obiettivo di questo paragrafo è l'analisi di alcune proprietà di questi sottoinsiemi dello spazio $\mathbb{R}^2$.
Consideriamo, dunque, un insieme $A \subseteq \mathbb{R}^2$. Diremo che esso è limitato se è contenuto in un disco aperto di centro l'origine (e raggio R) cioè

$$(x, y) \in D(0, R), \qquad \forall (x, y) \in A,$$

o, in maniera equivalente,

$$\|\boldsymbol{v}\| < R, \qquad \forall \boldsymbol{v} = (x, y) \in A.$$

In realtà, per dire che un insieme è limitato basta provare che esiste un intorno circolare qualsiasi che lo contiene senza necessariamente avere come centro l'origine degli assi (si prendano, come esempi, le Figure 2.5a e 2.5b).

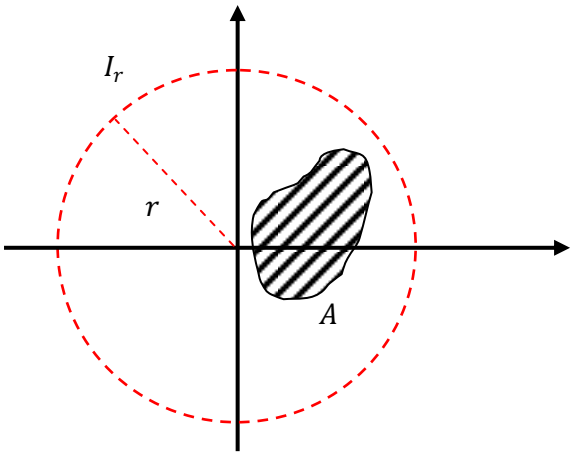

Figura 2.5a. L'insieme A è limitato.

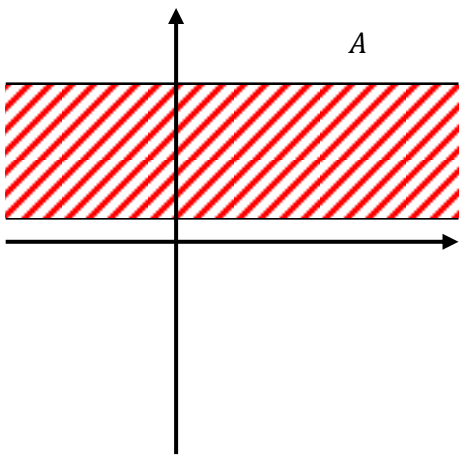

Figura 2.5b. L'insieme A non è limitato.

Tuttavia la definizione appena enunciata resta comunque valida poiché un qualunque intorno circolare di raggio r è contenuto in un disco aperto di centro l'origine per R opportuno.
Se, infatti, $A \subseteq I_r(x_0, y_0)$, posto $\boldsymbol{v_0} = (x_0, y_0)$, allora $A \subseteq D(0, R)$ nel quale $R > r + \|\boldsymbol{v}\|$.
Un aperto $A \subseteq \mathbb{R}^2$ è connesso se non esistono due (o più) aperti disgiunti non vuoti di $\mathbb{R}^2$ tali che la loro unione sia l'insieme A. In altre parole non esistono $A_1, A_2 \subseteq \mathbb{R}^2$ aperti con $A_1 \neq \emptyset$, $A_2 \neq \emptyset$

$$\begin{cases} A_1 \cap A_2 = \emptyset \\ A_1 \cup A_2 = A \end{cases}$$

In caso contrario l'insieme A non è connesso.
Un sottoinsieme di $\mathbb{R}^2$ è detto dominio se è la chiusura di un aperto. Un dominio, dunque, è connesso se è la chiusura di un aperto connesso.

2.5 FUNZIONI DI DUE VARIABILI (REALI)

La trattazione delle funzioni di due variabili è differente rispetto a quella delle funzioni ad una variabile, tuttavia, più che nella forma, questa differenza sta nella sostanza.
Molte definizioni, infatti, possono adattarsi naturalmente a funzioni con due variabili ricordandosi della doppia dipendenza di f e del fatto che in $\mathbb{R}^2$ come *intervallo* si considera l'intorno circolare visto nei paragrafi precedenti.

2.5.1 Introduzione

Una funzione di due variabili reali definita su $A \subseteq \mathbb{R}^2$ è un'applicazione che ad ogni punto $(x, y) \in A$ associa un unico valore reale. Un esempio è la scrittura $f(x, y) = x^2 y$.
Possiamo notare che se fissiamo $y_0 \in \mathbb{R}$, la funzione $f(x, y_0)$ è una funzione di un'unica variabile poiché y_0 viene considerato come una costante. Un discorso analogo vale se fissiamo $x_0 \in \mathbb{R}$.
Vediamo di fare un esempio: sia f una funzione di due variabili reali definita nel modo seguente

$$f(x, y) = x^2 e^{x+y}.$$

In analogia alle funzioni di una variabile possiamo anche dire

$$f: \mathbb{R}^2 \to \mathbb{R} \qquad (x, y) \in \mathbb{R}^2 \longmapsto x^2 e^{x+y}$$

Ora, fissato per esempio $x_0 = 1$ la nostra funzione diventa

$$f(x_0, y) = 1^2 \cdot e^{1+y} = e^{y+1}.$$

Oppure, con $x_0 = -2$, abbiamo

$$f(x_0, y) = (-2)^2 e^{-2+y} = 4e^{y-2}.$$

Analogamente, fissato $y_0 = 0$ otteniamo la scrittura

$$f(x, y_0) = x^2 e^{x+0} = x^2 e^x.$$

O, anche, per $y_0 = -1$

$$f(x, y_0) = x^2 e^{x-1}.$$

La differenza con le funzioni dipendenti da una singola variabile sta, dunque, in questa doppia dipendenza. Un metodo intuitivo ma solo teorico per la comprensione delle caratteristiche di una funzione a due variabili potrebbe essere il seguente: fissato x_0 (o y_0), si studia $f(x_0, y)$ (o $f(x, y_0)$) per poi iterare il procedimento per tutti i possibili $x_0 \in \mathbb{R}$ (o $y_0 \in \mathbb{R}$).
L'idea è quella di ricondursi alle funzioni di una variabile reale ma è inattuabile sul piano pratico poiché occorrerebbe considerare tutte le *infinite* funzioni di y (o di x) ottenute fissando, di volta in volta, l'altra variabile come costante.
Per lo studio analitico delle funzioni a due variabili occorrerà percorrere strade differenti: un'analisi sintetica ed esaustiva sarà l'obiettivo dei prossimi paragrafi in vista della sezione di richiami di analisi complessa.

2.5.2 Limiti e continuità

Ci proponiamo di estendere la definizione di limite e quella di continuità nel caso di funzioni a due variabili. La differenza sostanziale rispetto al caso di funzioni di una variabile reale sta nel fatto che questa doppia dipendenza implica una concezione di intorno non come intervallo ma come disco aperto. La funzione, quindi, tende al limite dalle infinite direzioni imposte dalla doppia variabile e non dall'unica direzione con cui avevamo a che fare nelle funzioni ad una sola incognita (Figure 2.6a e 2.6b).

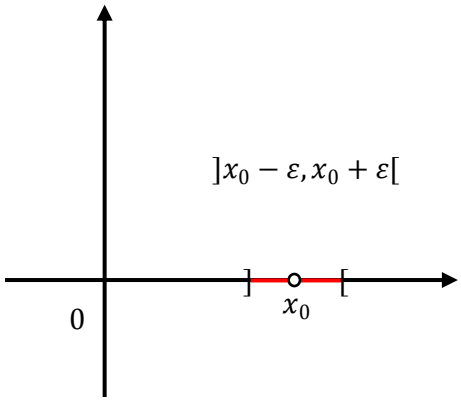

Figura 2.6a. Limite di una funzione di una variabile.

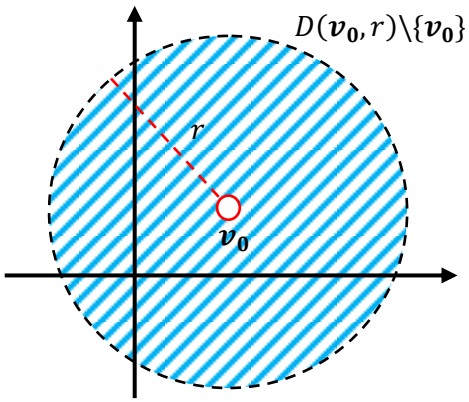

Figura 2.6b. Limite di una funzione a due variabili.

Le Figure 2.6a e 2.6b sono proprio l'esempio pratico della complessità del discorso nelle due variabili. La Figura 2.6a, infatti, mostra la situazione ad una variabile con x_0 punto limite e l'unica direzione – anche se con i due versi destro e sinistro (limite destro e sinistro) – con la quale si tende ad esso. Nella Figura 2.6b, è riportato l'intorno circolare del punto o vettore $\boldsymbol{v_0} = (x_0, y_0)$ di $\mathbb{R}^2$.
Come si può notare anche nella Figura 2.7, si può giungere al punto limite $\boldsymbol{v_0}$ da infinite direzioni, ognuna rappresentabile da una specifica retta tra tutte quelle del fascio di centro $\boldsymbol{v_0}$.

In analogia al caso di funzioni di una variabile, ogni direzione possiede un limite destro e sinistro e non è detto che questi siano uguali.
In quest'ambito terremo fede alla Figura 2.7 anche se, in realtà, si può giungere al punto limite $\boldsymbol{v_0}$ tramite una qualsiasi curva continua passante (o terminante) in (x_0, y_0).
Siano, dunque, $f: A \subseteq \mathbb{R}^2 \to \mathbb{R}^2$ e (x_0, y_0) un punto di accumulazione per A : vogliamo analizzare il limite per f che tende a (x_0, y_0) supponendo per ora che tale limite l sia finito, cioè $l \in \mathbb{R}$.

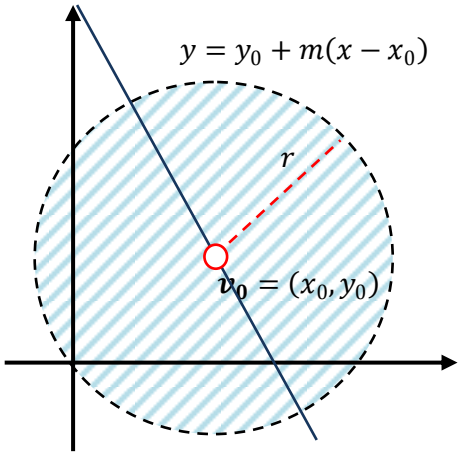

Figura 2.7. Analisi delle "infinite" direzioni del limite.

Diremo che $f(x, y)$ tende (o converge) a l per (x, y) che va a (x_0, y_0) se $\forall \varepsilon > 0, \exists \delta > 0$ tale che

$$|f(x, y) - l| < \varepsilon$$

per ogni $(x, y) \in A$, $(x, y) \neq (x_0, y_0)$ e $\|(x, y) - (x_0, y_0)\| < \delta$.
Scriveremo anche $f(x, y) \to l$ per $(x, y) \to (x_0, y_0)$ in analogia alle funzioni di un'unica variabile reale. Possiamo inoltre notare che la definizione di limite appena data è formalmente identica a quella di limite per funzioni di una variabile reale.
Ricordando la definizione di norma, l'ultima condizione si traduce però con

$$0 \neq \sqrt{(x - x_0)^2 + (y - y_0)^2} < \delta$$

I due casi di limite infinito – cioè $l = \pm\infty$ – si trattano in modo analogo fra loro; consideriamo di seguito solo il caso $l = +\infty$.
Diremo che $f(x, y)$ tende (o diverge) a $+\infty$ per (x, y) che tende a (x_0, y_0) se, qualunque sia $M > 0$, esiste $\delta > 0$ tale che

$$|f(x, y)| > M$$

$\forall (x, y) \in A$, $(x, y) \neq (x_0, y_0)$ e $\|(x, y) - (x_0, y_0)\| < \delta$.
In entrambi i casi, l'insieme dei punti soddisfacenti la condizione $\|(x, y) - (x_0, y_0)\| < \delta$ è proprio il disco aperto di centro (x_0, y_0) e raggio δ nella Figura 2.6b. Come detto in precedenza (Figura 2.7), invece di considerare tutti i punti del cerchio – privato del suo centro per definizione di limite ($(x, y) \neq (x_0, y_0)$) – ci si può limitare a considerare i punti (x, y) di tale insieme che appartengono ad una generica retta passante per il centro (x_0, y_0). In altre parole ci avviciniamo al punto limite tramite le infinite direzioni ognuna rappresentata da una generica retta tra quelle del fascio di centro (x_0, y_0) di equazione

$$y = y_0 + m(x - x_0).$$

Ogni $m \in \mathbb{R}$ coefficiente angolare individua una specifica retta.
Se ci limitiamo ai punti di questa retta, si ottiene che, se $f(x, y)$ ammette limite l – eventualmente infinito – per $(x, y) \to (x_0, y_0)$, posto $y = y_0 + m(x - x_0)$ risulta

$$|f(x, y_0 + m(x - x_0)| < \varepsilon$$

oppure

$$|f(x, y_0 + m(x - x_0)| > M$$

nel caso di limite infinito. Queste valgono per tutti i numeri reali $x \neq x_0$ e tali che

$$\sqrt{(x-x_0)^2+(y-y_0)^2} = \sqrt{(x-x_0)^2+m^2(y-y_0)^2} = \sqrt{1+m^2}\cdot\sqrt{(x-x_0)^2}$$
$$= \sqrt{1+m^2}\cdot|x-x_0| < \delta$$

Notiamo che, con la sostituzione $y = y_0 + m(x-x_0)$, la funzione $f\big(x, y_0 + m(x-x_0)\big)$ è una funzione di *una* variabile reale con parametro m. In base alla definizione di limite per funzioni di una variabile reale, questo significa che

$$\lim_{x\to x_0} f\big(x, y_0 + m(x-x_0)\big) = l.$$

Deduciamo che la condizione necessaria affinché esista il limite della funzione di due variabili

$$\lim_{(x,y)\to(x_0,y_0)} f(x,y) = l$$

è che il limite ottenuto con la sostituzione $y = y_0 + m(x-x_0)$ rimanga lo stesso al variare del coefficiente angolare m. In altre parole il limite non deve dipendere da m.
Vediamo di fare un esempio per avere un quadro più pratico riguardo alla questione.
Verifichiamo che

$$\lim_{(x,y)\to(0,0)} \frac{x^2y^2}{y} = 0$$

Applichiamo la sostituzione $y = y_0 + m(x-x_0)$, in questo caso $x_0 = 0$ e $y_0 = 0$.

$$\lim_{(x,y)\to(0,0)} \frac{x^2y^2}{y} = \lim_{x\to 0}\frac{x^2\cdot(m^2x^2)}{mx} = \lim_{x\to 0} mx^3 = 0$$

Il limite appena trovato vale $\forall m \in \mathbb{R}$ anche se resta da vedere, a parte, il caso $m = \infty$. Dalla teoria delle funzioni lineari ad una variabile, nel caso il cui il coefficiente angolare m sia infinito, vuol dire che la retta è del tipo $x = k$, nel nostro caso $x = x_0$. Sostituiamo, dunque, $x = 0$ e calcoliamo il limite nella *direzione* restante

$$\lim_{y\to 0} f(0,y) = \lim_{y\to 0} 0\cdot\frac{y^2}{y} = \lim_{y\to 0} 0\cdot y = 0$$

che è uguale a quello trovato nel caso precedente con $m \in \mathbb{R}$. Possiamo concludere che $f(x,y) \to 0$ per $(x,y) \to (0,0)$.
Tuttavia, come nel caso di funzioni di una variabile reale, non è sempre necessario calcolare il limite per via analitica e si può ricorrere ad altri metodi come le usuali maggiorazioni.
Verifichiamo, ad esempio, che

$$\lim_{(x,y)\to(0,0)} \frac{x^2}{\sqrt{x^2+y^2}} = 0$$

Possiamo applicare una serie di disuguaglianze

$$0 \leq \frac{x^2}{\sqrt{x^2+y^2}} \leq \frac{x^2+y^2}{\sqrt{x^2+y^2}} = \sqrt{x^2+y^2}$$

che tende a zero per $(x,y) \to (0,0)$.
Vediamo, ora, di estendere la definizione di continuità alle funzioni di due variabili.
Sia $f: A \subseteq \mathbb{R} \to \mathbb{R}$ una funzione di due variabili e (x_0, y_0) un punto (di accumulazione) per A. Si dice che la funzione $f(x,y)$ è continua in (x_0,y_0) se, per ogni $\varepsilon > 0$, esiste $\delta > 0$ tale che

$$|f(x,y) - f(x_0,y_0)| < \varepsilon$$

per ogni $(x,y) \in A$ con $\|(x,y) - (x_0,y_0)\| < \delta$.
In altre parole $f(x,y)$ è continua in (x_0,y_0) se

$$\lim_{(x,y)\to(x_0,y_0)} f(x,y) = f(x_0,y_0).$$

Per estensione diremo che $f(x,y)$ è continua in A se è continua $\forall(x_0,y_0) \in A$. Possiamo osservare che queste definizione di continuità estendono quelle delle funzioni di una variabile reale.

Vedremo, ora, che per le funzioni di due variabili reali valgono dei risultati analoghi a quelli delle funzioni ad una variabile reale.

Teorema

Siano $f(x,y)$ e $g(x,y)$ due funzioni continue definite su $A \subseteq \mathbb{R}^2$ e $(x_0,y_0) \in A$ un punto di accumulazione per entrambe.

- Se $0 \leq f(x,y) \leq g(x,y)$ in un intorno circolare centrato in (x_0,y_0) e $g(x,y) \to 0$ per $(x,y) \to (x_0,y_0)$ allora anche $f(x,y) \to 0$ per $(x,y) \to (x_0,y_0)$.
- Se $0 \geq f(x,y) \geq g(x,y)$ in un intorno circolare centrato in (x_0,y_0) e $g(x,y) \to 0$ per $(x,y) \to (x_0,y_0)$ allora anche $f(x,y) \to 0$ per $(x,y) \to (x_0,y_0)$.
- Se $f(x,y) \geq g(x,y)$ in un intorno circolare centrato in (x_0,y_0) e $g(x,y) \to +\infty$ allora anche $f(x,y) \to +\infty$ per $(x,y) \to (x_0,y_0)$.
- Se $f(x,y) \leq g(x,y)$ in un intorno circolare centrato in $(x_0,y_0) \to (x_0,y_0)$ e $g(x,y) \to -\infty$ allora $f(x,y) \to -\infty$ per $(x,y) \to (x_0,y_0)$.

Questo teorema ci consente di effettuare dei confronti tra funzioni nel calcolo dei limiti e giustifica il risultato dell'esercizio precedente.

Teorema (proprietà del limite)

Siano $f(x,y)$ e $g(x,y)$ funzioni continue definite su $A \subseteq \mathbb{R}^2$ e $(x_0,y_0) \in A$ un punto di accumulazione per entrambe. Allora:

- $\lim_{(x,y)\to(x_0,y_0)}\big(f(x,y) \pm g(x,y)\big) =$
  $\lim_{(x,y)\to(x_0,y_0)} f(x,y) \pm \lim_{(x,y)\to(x_0,y_0)} g(x,y)$;
- $\lim_{(x,y)\to(x_0,y_0)}\big(f(x,y) \cdot g(x,y)\big) = \lim_{(x,y)\to(x_0,y_0)} f(x,y) \cdot \lim_{(x,y)\to(x_0,y_0)} g(x,y)$;
- $\lim_{(x,y)\to(x_0,y_0)}\big(c \cdot f(x,y)\big) = c \cdot \lim_{(x,y)\to(x_0,y_0)} f(x,y)$ per $c \in \mathbb{R}$.

Tutte queste proprietà valgono nei casi in cui i limiti esistano e non siano forme indeterminate.

Teorema

Siano $f(x,y)$ e $g(x,y)$ due funzioni continue definite su $A \subseteq \mathbb{R}^2$. Allora:

- $f(x,y) \pm g(x,y)$ è continua in A,
- $c \cdot f(x,y)$ è continua in A, $\forall c \in \mathbb{R}$,
- $f(x,y) \cdot g(x,y)$ è continua in A,
- $\frac{f(x,y)}{g(x,y)}$ è continua in A tranne nei punti in cui $g(x,y) = 0$,
- $f(x,y)^{g(x,y)}$ è continua in A (tranne in caso di forme indeterminate).

Teorema (Weierstrass)

Siano $C \subseteq \mathbb{R}^2$ un insieme chiuso e limitato e $f: C \to \mathbb{R}^2$ continua. Allora f assume massimo e minimo su C, cioè esistono due punti (x_1,y_1) e (x_2,y_2) di C tali che

$$f(x_1,y_1) \leq f(x,y) \leq f(x_2,y_2) \qquad \forall(x,y) \in C.$$

Teorema (Cantor)
Siano $C \subseteq \mathbb{R}^2$ un insieme chiuso e limitato e $f: C \to \mathbb{R}^2$ continua. Allora f è uniformemente continua su C, cioè per ogni $\varepsilon > 0$ esiste $\delta > 0$ tale che

$$|f(x_1, y_1) - f(x_2, y_2)| < \varepsilon$$

per ogni $(x_1, y_1), (x_2, y_2) \in C$, con $\|(x_1, y_1) - (x_2, y_2)\| < \delta$.

Teorema (esistenza dei valori intermedi)
Siano D un dominio connesso e limitato di $\mathbb{R}^2$ e $f(x, y)$ una funzione continua definita su D. Allora f assume tutti i valori compresi tra il minimo e il massimo di f su D.

2.5.3 Derivate parziali

Siano $A \subseteq \mathbb{R}^2$ un aperto, $f: A \to \mathbb{R}$ una funzione di due variabili e $(x, y) \in A$ un punto fissato. Se il seguente limite

$$\lim_{h \to 0} \frac{f(x + h, y) - f(x, y)}{h}$$

esiste ed è finito, esso è la derivata parziale di f rispetto ad x nel punto (x, y). Tale derivata parziale di f rispetto ad x si indica con uno dei seguenti simboli

$$\frac{\partial f}{\partial x}, \qquad \frac{\partial}{\partial x} f, \qquad f_x, \qquad f_x(x, y).$$

Tra questi, nei testi di matematica i più usati sono di gran lunga i primi due anche se il terzo è comodo come abbreviazione.
Analogamente, la derivata parziale di f rispetto ad y nel punto (x, y) è il limite

$$\lim_{k \to 0} \frac{f(x, y + k) - f(x, y)}{k}$$

ammesso che tale limite esista e sia finito. Tale derivata è indicata con uno dei seguenti simboli

$$\frac{\partial f}{\partial y}, \qquad \frac{\partial}{\partial y} f, \qquad f_y, \qquad f_y(x, y).$$

Se in un punto (x, y) esistono entrambe le derivate parziali f_x e f_y diremo che la funzione f è derivabile in (x, y). Inoltre, se f è derivabile in ogni punto $(x, y) \in A$ si dice che f è derivabile in A. Inoltre, in analogia al caso di funzioni di una variabile reale, se una funzione è derivabile in A e le sue derivate parziali sono funzioni continue, essa è di classe C^1 in A.
Possiamo notare dalle definizioni che le due derivate parziali di f rispetto a x o y si calcolano considerando l'altra variabile – rispettivamente la y o x – fissata, quindi costante. Nel calcolo di una derivata parziale, entra in gioco una sola variabile reale alla volta: ci si può dunque servire delle usuali regole di derivazione valide per le funzioni di una variabile reale.
Consideriamo, ad esempio la seguente funzione

$$f(x, y) = x^3 y + e^x - y^2.$$

Allora:

- $f_x = 3x^2 y + e^x$,
- $f_y = x^3 - 2y$.

Facciamo un altro esempio: consideriamo la seguente funzione

$$f(x,y) = y^2 \cos(x^2).$$

Allora:

- $f_x = 2xy^2 \sin(x^2)$,
- $f_y = 2y \cos(x^2)$.

Sul piano pratico, dunque, il calcolo di una derivata parziale non è molto differente da quello usuale per le derivate di una funzione di una variabile reale. Con abuso di termini possiamo concludere che nel calcolo di una derivata parziale la variabile non interessata dalla derivazione è considerata alla stregua di un parametro.

A tal proposito, consideriamo la seguente funzione di *una* variabile reale

$$f(x) = (x-k)e^{x+k}$$

Il calcolo della sua derivata è quello usuale, ricordandoci del parametro $k \in \mathbb{R}$ considerato alla stregua di una costante.

$$f'(x) = e^{x+k} + e^{x+k}(x-k) = e^{x+k}[1+x-k]$$

Ora, invece, consideriamo la seguente funzione di *due* variabili reali

$$f(x,y) = (x-y)e^{x+y}$$

Se calcoliamo la derivata parziale di f rispetto a y otteniamo

$$f_x(x,y) = e^{x+y} + e^{x+y}(x-y) = e^{x+y}[1+x-y]$$

formalmente identica – a parte y al posto di k – a quella vista nell'esempio precedente.

2.5.4 Derivate successive

Sia $f(x,y)$ una funzione di due variabili reali definita in $A \subseteq \mathbb{R}^2$. Supponiamo che essa sia di classe C^1 in A, cioè che esistano le due derivate parziali f_x e f_y e che siano funzioni continue in A. Esse, a loro volta, possono essere funzioni derivabili: se ciò si verifica le loro derivate (parziali)

$$\frac{\partial}{\partial x} f_x, \qquad \frac{\partial}{\partial y} f_y, \qquad \frac{\partial}{\partial x} f_y, \qquad \frac{\partial}{\partial y} f_y$$

sono dette derivate seconde della funzione $f(x,y)$. Se esistono diremo che f è derivabile 2 volte in A. Se queste derivate sono continue f è chiamata di classe C^2.

Le quattro derivate parziali seconde, si classificano in due gruppi.

- Derivate seconde *pure*, quelle nelle quali si deriva due volte f rispetto alla stessa variabile, dunque $\frac{\partial}{\partial x} f_x, \frac{\partial}{\partial y} f_y$. Si indicano con f_{xx} , f_{yy} oppure con $\frac{\partial^2 f}{\partial x^2}$ e $\frac{\partial^2 f}{\partial y^2}$ rispettivamente.
- Derivate seconde *miste*, quelle nelle quali si deriva prima rispetto ad una variabile poi rispetto all'altra e sono $\frac{\partial}{\partial y} f_x, \frac{\partial}{\partial x} f_y$. Si indicano con f_{xy}, f_{yx} oppure con $\frac{\partial^2 f}{\partial x \partial y}$ e $\frac{\partial^2 f}{\partial y \partial x}$.

Spesso le quattro derivate parziali seconde di f si dispongono sotto forma di matrice 2×2

$$D^2(f) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix}.$$

La matrice delle derivate seconde è detta matrice Hessiana e si denota con il simbolo $D^2(f)$ oppure $H(f)$ anche se quest'ultimo è utilizzato più raramente.
A titolo di esempio, consideriamo la seguente funzione

$$f(x,y) = x^2 e^{2y}.$$

Allora:

- $f_x = 2xe^{2y}, f_y = 2x^2e^{2y}$,
- $f_{xx} = 2e^{2y}, f_{xy} = 4xe^{2y}, f_{yx} = 4xe^{2y}, f_{yy} = 4x^2e^{2y}$.

Disponendo le quattro derivate seconde nella matrice Hessiana

$$D^2(f) = \begin{pmatrix} 2e^{2y} & 4xe^{2y} \\ 4xe^{2y} & 4x^2e^{2y} \end{pmatrix}$$

Da notare che le derivate seconde miste sono uguali: in generale vale il seguente risultato.

Teorema (Schwarz)
Siano $A \subseteq \mathbb{R}^2$ aperto, $(x_0, y_0) \in A$ e $f(x,y)$ una funzione derivabile due volte in A. Se le derivate seconde miste f_{xy} e f_{yx} sono continue in (x_0, y_0) allora $f_{xy}(x_0, y_0) = f_{yx}(x_0, y_0)$.

Il teorema appena visto può essere esteso nel seguente modo: se $f \in C^2$ allora le derivate parziali seconde miste sono uguali. In alcuni testi si usa dire che le derivate seconde miste *commutano*: è un termine riferito agli indici che si possono scambiare proprio perché esse sono uguali.
E' naturale, a questo punto, andare oltre il secondo ordine e, se la funzione è sufficientemente regolare, definire le derivate successive a quelle parziali seconde. In modo analogo avremo le derivate parziali terze, quarte,... e così via.
Diremo che $f(x,y)$, definita su $A \subseteq R^2$ aperto è di classe C^n in $(x_0, y_0) \in A$, con $n \geq 1$ e intero, se le derivate parziali $n-$esime esistono e sono continue in (x_0, y_0). Possiamo estendere in modo naturale tale definizione concludendo che f è di classe C^n in A se è derivabile con continuità fino al $n-$esimo ordine e le derivate parziali sono continue in A.
L'unico inconveniente è che le derivate di qualsiasi ordine sono parziali e, dunque, alle quattro derivate seconde seguiranno le otto derivate parziali terze, le sedici quarte e così via.
Riprendiamo l'esempio precedente

$$f(x,y) = x^2 e^{2y}.$$

Avremo:

- $f_x = 2xe^{2y}, f_y = 2x^2e^{2y}$,
- $f_{xx} = 2e^{2y}, f_{xy} = 4xe^{2y}, f_{yx} = 4xe^{2y}, f_{yy} = 4x^2e^{2y}$,
- $f_{xxx} = 0, f_{xxy} = 4e^{2y}, f_{xyx} = 4e^{2y}, f_{xyy} = 8xe^{2y}, f_{yxx} = 4e^{2y}, f_{yxy} = 8xe^{2y}, f_{yyx} = 8xe^{2y}, f_{yyy} = 8x^2e^{2y}$.

La funzione dell'esempio è di classe C^∞ e si potrebbe andare avanti ad oltranza nel calcolo delle derivate complicando sempre di più la situazione passando da un ordine al successivo.
Si può notare che $f_{xxy} = f_{xyx}$ in accordo al teorema di Schwarz poiché f_x è una funzione di due variabili di classe C^2. In fondo f_{xxy} e f_{xyx} non sono altro che le derivate parziali seconde

miste della funzione f_x. Analogamente $f_{yxy} = f_{yyx}$ poiché $f_y \in C^2$. Inoltre $f_{xyx} = f_{yxx}$ e $f_{xyy} = f_{yxy}$ poiché, in entrambi i casi, si tratta di derivare le derivate parziali miste del secondo ordine (uguali per il teorema di Schwarz) rispetto alla stessa variabile.
L'osservazione appena vista permette di trarre un'importante conclusione. Sebbene $f(x,y)$ abbia – ammesso che esistano – 2^n derivate parziali di ordine n ($n \geq 1$ e intero), il teorema di Schwarz semplifica parecchio questo discorso nel caso di funzioni di classe C^n riducendole a $n+1$. Infatti, per una $f(x,y)$ di classe C^n le derivate parziali miste di ordine n commutano a due a due e si riducono alle seguenti

$$\frac{\partial^n f}{\partial x^n}, \quad \frac{\partial^n f}{\partial x^{n-1}\partial y}, \quad \frac{\partial^n f}{\partial x^{n-2}\partial y}, \quad \dots, \quad \frac{\partial^n f}{\partial x \partial y^{n-1}}, \quad \frac{\partial^n f}{\partial y^n}$$

Una precisazione sulla notazione appena utilizzata.
Per quanto riguarda le derivate successive, la scrittura compatta f_{xxy} è poco utilizzata per le derivate terze e per nulla utilizzata in quelle successive. Si preferisce come unica rappresentazione quella generale $\frac{\partial^3 f}{\partial x^2 \partial y}$ in luogo di f_{xxy}. In essa l'indice sopra al simbolo della derivata parziale ("∂") indica quante volte si deriva mentre al *denominatore* compaiono le variabili di derivazione nell'ordine in cui si effettua la derivazione stessa.
Per esempio, la scrittura

$$\frac{\partial^4 f}{\partial x^2 \partial y \partial x}$$

indica la derivata parziale quarta di f ottenuta derivando 2 volte rispetto a x, poi una rispetto a y e infine un'ultima volta di nuovo rispetto a x. Da notare che si usa ∂x^2 al posto di $\partial x \partial x$ e, in generale, ∂x^n in luogo di $\partial x \partial x \cdots \partial x$ ripetuto n volte.

Teorema
Siano $f(x,y), g(x,y)$ due funzioni di due variabili di classe C^n in $A \subseteq \mathbb{R}^2$. Allora:
- $f(x,y) \pm g(x,y)$ è di classe C^n in A,
- $c \cdot f(x,y)$ è di classe C^n in A, per ogni $c \in \mathbb{R}$,
- $f(x,y) \cdot g(x,y)$ è di classe C^n in A,
- $f(x,y)/g(x,y)$ è C^n in A ad eccezione (eventualmente) dei punti nei quali $g(x,y) = 0$,
- $f(x,y)^{g(x,y)}$ è C^n, salvo casi in cui si ha a che fare con forme indeterminate.

In generale la composizione tra $f(x,y)$ e/o $g(x,y)$ e una qualsiasi funzione di classe C^n – in particolar modo con le funzioni elementari (polinomiali, trigonometriche...) – dà come risultato una funzione di classe C^n.

2.5.5 Gradiente e punti critici

Sia $f(x,y)$ una funzione derivabile in un punto $(x_0, y_0) \in A$ con $A \subseteq \mathbb{R}^2$ aperto. Il gradiente di f in (x_0, y_0) è il vettore $\nabla f \in \mathbb{R}^2$ le cui componenti sono le derivate parziali di f calcolate in (x_0, y_0)

$$\nabla f(x_0, y_0) = \left(\frac{\partial f}{\partial x}(x_0, y_0), \frac{\partial f}{\partial y}(x_0, y_0)\right) = \left(f_x(x_0, y_0), f_y(x_0, y_0)\right).$$

In generale, se $f(x, y)$ è derivabile in A, al variare di $(x, y) \in A$ il gradiente di f in A è il vettore ∇f con componenti le due derivate parziali

$$\nabla f(x, y) = \left(\frac{\partial f}{\partial x}(x, y), \frac{\partial f}{\partial y}(x, y)\right) = \left(f_x(x, y), f_y(x, y)\right).$$

Per indicare il gradiente, sono generalmente utilizzati – in modo equivalente – i seguenti simboli

$$\nabla f, \qquad Df(x, y), \qquad \mathrm{grad}(f).$$

Sia $f(x, y)$ definita su $A \subseteq \mathbb{R}^2$. Un punto $(x_0, y_0) \in A$ si dice di massimo relativo per f su A se esiste un intorno circolare $I_\delta(x_0, y_0)$ centrato in (x_0, y_0) di raggio $\delta > 0$ tale che

$$f(x_0, y_0) \geq f(x, y), \qquad \forall (x, y) \in I_\delta(x_0, y_0).$$

Analogamente un punto $(x_0, y_0) \in A$ è di minimo relativo per f su A se esiste un intorno circolare $I_\delta(x_0, y_0)$ centrato in (x_0, y_0) di raggio $\delta > 0$ tale che

$$f(x_0, y_0) \leq f(x, y), \qquad \forall (x, y) \in I_\delta(x_0, y_0).$$

Condizione necessaria (primo ordine)

Se una funzione f definita su $A \subseteq \mathbb{R}^2$ è derivabile in (x_0, y_0) punto di massimo o di minimo relativo per f, allora il gradiente di f si annulla in (x_0, y_0).

Questa condizione viene detta anche "condizione del primo ordine" poiché riguarda le derivate parziali prime di f.

Diremo che $(x_0, y_0) \in A$ è un punto critico per f se in esso si annulla il gradiente. La condizione necessaria, infatti, non è anche sufficiente poiché se $\nabla f(x_0, y_0) = 0$ non è detto che il punto critico sia di minimo o di massimo. Può accadere che un punto critico sia né di massimo né di minimo: in questo caso è detto anche punto di sella.

Il motivo della nomenclatura "punto di sella" è dovuto al fatto che il grafico della funzione con gradiente nullo in un punto né di massimo né di minimo assume una forma curiosa che ricorda da vicino una sella. I grafici delle funzioni a due variabili e delle funzioni di variabile complessa saranno trattati in una sezione a parte.

Nel caso in cui f sia derivabile più di una volta, valgono altre condizioni per stabilire la natura dei punti critici. Ricordiamo la matrice Hessiana di una funzione f

$$D^2(f) = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}$$

Il suo determinante è

$$Hf(x, y) = \begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix} = f_{xx} f_{yy} - f_{xy} f_{yx}.$$

A tale scopo, possiamo definire la seguente condizione sufficiente del secondo ordine.

Condizione sufficiente (secondo ordine)

Siano $A \subseteq \mathbb{R}^2$, $(x_0, y_0) \in A$ e f di classe C^2 definita su A. Se vale

$$\begin{cases} \nabla f(x_0, y_0) = 0 \\ f_{xx}(x_0, y_0) > 0, \qquad f_{yy}(x_0, y_0) > 0, \\ Hf(x_0, y_0) > 0 \end{cases}$$

allora (x_0, y_0) è un punto di minimo relativo per f su A. Se, invece

$$\begin{cases} \nabla f(x_0, y_0) = 0 \\ f_{xx}(x_0, y_0) < 0, \qquad f_{yy}(x_0, y_0) < 0, \\ Hf(x_0, y_0) > 0 \end{cases}$$

allora (x_0, y_0) è un punto di massimo relativo per f su A.
Negli altri casi (x_0, y_0) è un punto critico per f ma non è né di massimo né di minimo.

Questa condizione è detta del secondo ordine poiché riguarda le derivate parziali seconde di f. Vediamo, ora, di fare un esempio: consideriamo la funzione $f(x, y) = x^2 + y^2$, definita su $\mathbb{R}^2$. E' facile notare che $f(x, y) \geq 0, \forall (x, y) \in \mathbb{R}^2$ e che, nello specifico,

- $f(x, y) > 0, \forall (x, y) \in \mathbb{R}^2 \backslash \{(0,0)\}$
- $f(x, y) = 0$ per $(x, y) = (0,0)$,

dunque il punto $(0,0)$ è di minimo per f. In questo caso $(0,0)$ è di minimo globale poiché

$$\forall (x, y) \in \mathbb{R}^2 \backslash \{(0,0)\}, \qquad f(x, y) > f(0,0).$$

Vediamo di calcolare il gradiente nel punto $(0,0)$.

$$\nabla f = (2x, 2y), \qquad \nabla f(0,0) = (0,0),$$

da cui si può dedurre che l'origine degli assi è un punto critico per f. Passiamo all'Hessiano.

$$D^2(f) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

del quale il determinante è

$$Hf(0,0) = \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} = 2 \cdot 2 - 0 \cdot 0 = 4 > 0.$$

Da notare che in questo caso specifico, la matrice Hessiana (o Hessiano di f) non dipende dai punti considerati ed è ad elementi costanti. In genere, invece, gli elementi della matrice Hessiana sono delle funzioni di due variabili che dipendono dalla scelta dei punti (x, y).
Poiché $f_{xx}(0,0) = 2 > 0$, $f_{yy}(0,0) = 2 > 0$ e $Hf(0,0) = 4 > 0$ possiamo dedurre dalla condizione sufficiente che il punto $(0,0)$ è di minimo per f.
Abbiamo voluto proporre un esempio abbastanza semplice: in realtà quanto detto vale per qualsiasi funzione di due variabili di classe C^2. La scelta di $f(x, y) = x^2 + y^2$ è motivata dal fatto che si poteva dedurre a priori che l'origine era un punto di minimo per poi mostrare come applicare la condizione sufficiente.

2.6 CURVE NEL PIANO

2.6.1 Introduzione

C'è una particolare classe di funzioni composte, dette *curve* del piano.

In apparenza sembrano essere funzioni in due variabili del tipo appena visto ma, in realtà, entrambe le variabili sono a loro volta dipendenti da uno stesso parametro reale che le rende più simili a funzioni di un'unica variabile piuttosto che di due.
Nella prossima sezione, serviranno da background per lo studio degli integrali curvilinei che non sono altro che particolari tipi di integrali calcolati lungo le curve stesse invece che lungo intervalli.
Per quanto riguarda gli integrali curvilinei, essi saranno trattati a fondo direttamente nell'ambito dell'analisi complessa mentre in questa sezione forniremo una breve introduzione sulle curve che verrà ripresa e riadattata nel campo complesso per poi passare all'integrazione complessa.

2.6.2 Curve nel piano (in $\mathbb{R}^2$)

Consideriamo due funzioni $x(t), y(t)$ di variabile reale definite al variare di t in un intervallo $I \subseteq \mathbb{R}$. L'applicazione γ che ad ogni $t \in I$ associa il punto

$$\big(x(t), y(t)\big) \in \mathbb{R}^2$$

è una curva del piano.
Sebbene possa sembrare una funzione di due variabili reali – ognuna dipendente da un parametro t – essa è, come già detto, molto più affine ad una funzione (composta) di un'unica variabile reale. La sua definizione formale è la seguente: diremo che una curva del piano è una funzione

$$\gamma: I \subseteq \mathbb{R} \to \mathbb{R}^2, \qquad t \in I \longmapsto \gamma(t) = \big(x(t), y(t)\big) \in \mathbb{R}^2$$

Essa viene anche detta curva parametrica proprio per sottolineare il fatto che ogni sua componente dipende dal parametro reale t. Un modo alternativo per la sua rappresentazione è rappresentare singolarmente le sue componenti in un sistema

$$\begin{cases} x(t) \\ y(t) \end{cases}$$

Vediamo di fare degli esempi.
Consideriamo la seguente curva parametrica

$$\begin{cases} x(t) = t + 2 \\ y(t) = t^2 \end{cases}$$

Al variare di $t \in [0,1]$ essa rappresenta il luogo dei punti nel piano con ascissa $t + 2$ e ordinata t^2; in questo caso possiamo operare qualche calcolo

$$x = t + 2, \qquad x - 2 = t$$

per poi sostituirlo nella componente y ottenendo

$$y = (x - 2)^2 = x^2 - 4x + 4.$$

dunque il punto di coordinate $\big(x(t), y(t)\big)$ appartiene all'equazione $y = x^2 - 4x + 4$ con $x \in [2,3]$.
Consideriamo, ora, la seguente curva del piano

$$\begin{cases} x(t) = \cos t \\ y(t) = \sin t \end{cases}$$

con $t \in [0,2\pi]$. Otteniamo

$$\big(x(t)\big)^2 + \big(y(t)\big)^2 = \cos^2 t + \sin^2 t = 1$$

pertanto il punto di coordinate $\big(x(t), y(t)\big)$ appartiene alla circonferenza di equazione $x^2 + y^2 = 1$, cioè la circonferenza con centro nell'origine e raggio 1, cioè la circonferenza goniometrica.

Le curve parametriche sono utilizzate spesso in fisica, nell'ambito del moto. Esse rappresentano la posizione di un corpo al variare del tempo t nell'intervallo considerato e, dunque, la traiettoria.

Dal punto di vista matematico, il ragionamento potrebbe essere più ampio.

- *In linea teorica*, a partire da ogni curva parametrica, si può sperare di trovare una corrispettiva funzione di variabile reale. Il modo più semplice è quello di esplicitare il parametro t in una delle due componenti per poi sostituirlo nell'altra anche se non sempre si rivela efficace.
  Ovviamente questo non è sempre agevole e, in genere, ci si serve della scrittura parametrica senza cercare di esplicitare l'equazione della curva rispetto ad una variabile.
- Viceversa, ad ogni funzione di una variabile reale, corrisponde (almeno) una scrittura in forma parametrica. Il modo più semplice è quello di porre $x = t$ e $y = f(t)$
  $$y = f(x) \quad \Longrightarrow \quad \big(x(t), y(t)\big), \qquad x(t) = t, \qquad y(t) = f(t),$$
  per esempio, a partire dalla funzione $y = x^3 + x$, otteniamo $(t, t^3 + t)$.

Indichiamo con A un insieme aperto di $\mathbb{R}^2$ che contiene i punti $\big(x(t), y(t)\big)$ al variare di $t \in I$. Sia $f(x, y)$ una funzione di due variabili reali definita in A. Possiamo considerare la funzione composta

$$F(t) = f\big(x(t), y(t)\big), \qquad t \in I$$

Se indichiamo con $\gamma(t)$ la corrispettiva curva, $\gamma(t) = \big(x(t), y(t)\big)$, la funzione composta

$$F(t) = f\big(x(t), y(t)\big), \qquad t \in I$$

si indica anche, in forma compatta, con

$$F(t) = f\big(\gamma(t)\big), \qquad t \in I$$

Nel complesso, essa *non* è una funzione di due variabili reali, ma di una sola variabile e ricorrerà spesso nel calcolo degli integrali curvilinei. La sua derivata rispetto a t è

$$F'(t) = f_x\big(x(t), y(t)\big)x'(t) + f_y\big(x(t), y(t)\big)y'(t)$$

o, in forma compatta

$$F'(t) = f\big(\gamma(t)\big)\gamma'(t)$$

Quest'ultima scrittura necessita di un'ulteriore interpretazione. Prima, però, è utile considerare qualche altra definizione e proprietà circa le curve nel piano. A tal proposito consideriamo la curva $\gamma(t) = \big(x(t), y(t)\big)$ definita per $t \in I \subseteq \mathbb{R}$. Salvo indicazioni contrarie in loco, $I = [a, b]$.

- $\gamma(t)$ è una curva continua se le sue componenti sono entrambe funzioni continue $\forall t \in I$. Le curve che considereremo saranno sempre continue salvo eventuali rettifiche.
- $\gamma(t)$ è derivabile se le sue componenti sono entrambe funzioni derivabili. La derivata di $\gamma(t)$ è $\gamma'(t) = \big(x'(t), y'(t)\big)$.
- $\gamma(t)$ è differenziabile – o *liscia* – se le sue componenti sono entrambe differenziabili al variare di t.

- In generale $\gamma(t)$ è una curva di classe C^n se le sue componenti sono entrambe funzioni di classe C^n in I.
- $\gamma(t)$ è una curva regolare se è (almeno) di classe C^1 e $\gamma'(t) \neq 0, \forall t \in I$.
  In questo caso occorre fare attenzione: $\gamma'(t) = 0$ equivale a richiedere che si annullino contemporaneamente entrambe le derivate delle componenti.
- $\gamma(t)$ si dice regolare a tratti se esiste una partizione di $[a, b]$ t.c. γ è regolare in ciascun intervallo della partizione. In altre parole $\gamma(t)$ è l'unione di curve regolari.
- $\gamma(t)$ si dice chiusa se $\gamma(a) = \gamma(b)$.
- γ si dice semplice se non ha auto-intersezioni, cioè se $\gamma(t_1) \neq \gamma(t_2)$, $\forall t_1, t_2 \in [a, b]$ con $t_1 \neq t_2$ (escludendo, al più, $t_1 = t_2 = a = b$).

Torniamo, dunque, alla definizione di derivata per la funzione composta $F(t)$ la cui scrittura è, a questo punto, giustificata dal modo di derivare una curva componente per componente. Tuttavia, dal punto di vista formale, la rappresentazione compatta di $F'(t)$ trova un fondamento nel considerare $\gamma(t)$ come il vettore di componenti $\big(x(t), y(t)\big)$

$$F'(t) =< \nabla f, \gamma(t) >= f_x\big(x(t), y(t)\big)x'(t) + f_y\big(x(t), y(t)\big)y'(t).$$

La forma compatta è una rappresentazione teorica che trova un utilizzo più concreto nel campo complesso nel quale ci sono curve indicabili mediante una rappresentazione semplice senza esplicitare le componenti.

2.6.3 Spazi semplicemente connessi

Introdurremo brevemente la nozione di spazio semplicemente connesso per sottoinsiemi di $\mathbb{R}^2$.

Diremo che $A \subseteq \mathbb{R}^2$ è semplicemente connesso se valgono le seguenti condizioni:

- A è connesso;
- $\forall \gamma$ curva chiusa contenuta in A si ha che la regione all'interno di essa è tutta contenuta in A.

Consideriamo, ad esempio, le Figure 2.8a e 2.8b.

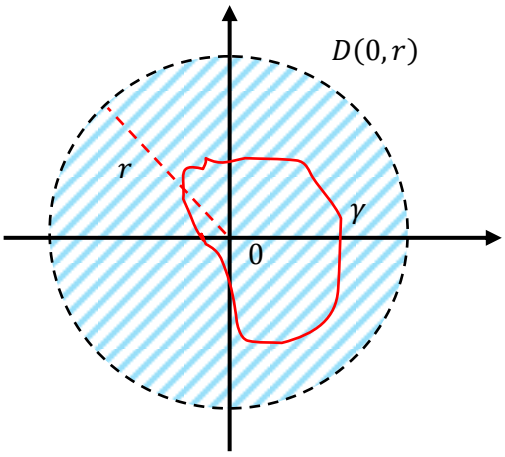

Figura 2.8a. Aperto semplicemente connesso.

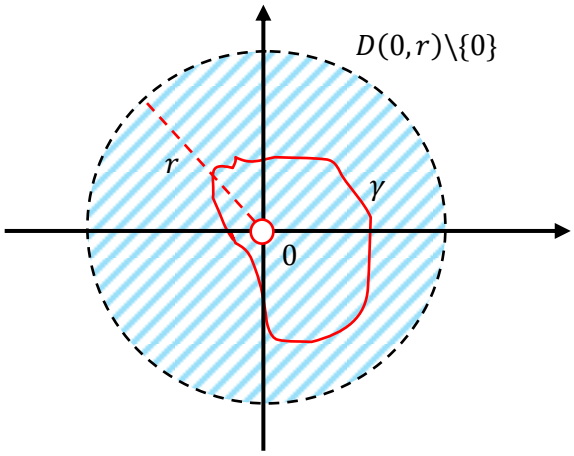

Figura 2.8b. Aperto non semplicemente connesso.

La Figura 2.8a mostra un aperto semplicemente connesso. Qualsiasi curva chiusa al suo interno – nella figura ne riportiamo una sola come esempio – è tale che la regione che racchiude è tutta contenuta in A. Nella Figura 2.8b, invece, l'interno della curva chiusa γ non è contenuto in A poiché l'origine non fa parte di A anche se appartiene all'interno di γ.

3. RICHIAMI DI ANALISI COMPLESSA

In questa sezione si richiameranno concetti importanti di Analisi Complessa; questi sono necessari poiché l'ipotesi di Riemann parla di una funzione ζ di variabile complessa. L'obiettivo, analogamente alle altre sezioni di richiamo, è quello di fornire gli strumenti necessari per comprendere gli argomenti che verranno esposti in seguito.
Questa sezione si propone dunque di essere un contenitore esaustivo ma non sovrabbondante di conoscenze fondamentali: in questo senso, analogamente alle altre sezioni, verranno omesse le dimostrazioni dei teoremi ed i temi che non sono necessari per la comprensione di questa tesi.

3.1 I NUMERI COMPLESSI

In questa sottosezione introdurremo l'insieme dei numeri complessi e tratteremo le loro proprietà. Vedremo in particolare come essi si possono rappresentare in vari modi tra cui uno "geometrico" – in realtà "trigonometrico" – molto utile, ad esempio, quando si tratta di calcolare potenze e radici n −esime.

3.1.1 Il campo complesso

Consideriamo l'insieme $\mathbb{C} = \{(x, y): x, y \in \mathbb{R}\}$. Di ogni sua coppia:

- x è detta parte reale,
- y è detta parte immaginaria.

$\mathbb{C}$ viene chiamato "insieme dei numeri complessi". Si può notare che la definizione data è molto simile a quella dell'insieme $\mathbb{R}^2$.
Nell'insieme $\mathbb{C}$ si definiscono:

- l'uguaglianza, ponendo $(x, y) = (x', y') \Leftrightarrow x = x', y = y'$;
- l'addizione, ponendo $(x, y) + (x', y') = (x + x', y + y')$;
- la moltiplicazione, ponendo $(x, y)(x', y') = (xx' - yy', xy' + yx')$.

Per quanto riguarda le operazioni definite nell'insieme $\mathbb{C}$ valgono le seguenti proprietà:

- l'addizione gode delle proprietà commutativa, associativa, c'è l'elemento neutro che è $(0,0)$ e di ogni (x, y) c'è l'opposto $(-x, -y)$;
- la moltiplicazione gode delle proprietà commutativa, associativa, distributiva rispetto alla somma, c'è poi un elemento neutro che è $(1,0)$;
- per ogni numero complesso $(x, y) \neq (0,0)$ esiste un inverso (lo vedremo a breve).

Di fatto, queste proprietà rendono l'insieme $\mathbb{C}$ con le sue operazioni un campo (vedere ([1] §1.3) per una trattazione più dettagliata).

Consideriamo, ora, la funzione $\varphi: \mathbb{R} \to \mathbb{C}$ tale che $\varphi(x) = (x, 0)$ con $x \in \mathbb{R}$ e $(x, 0) \in \mathbb{C}$.
Valgono le seguenti proprietà:

- φ è iniettiva, in altre parole $x \neq x'$ implica $\varphi(x) \neq \varphi(x')$ poiché $(x, 0) \neq (x', 0)$;
- φ rispetta le operazioni di addizione e moltiplicazione, nel senso che $\varphi(x) + \varphi(x') = \varphi(x + x')$ e $\varphi(x)\varphi(x') = \varphi(xx')$, $\forall x, x', \in \mathbb{R}$.

In altre parole φ è un omomorfismo iniettivo del campo reale in quello complesso, che identifica $\mathbb{R}$ col sottocampo di $\mathbb{C}$ formato dai numeri complessi della forma $(x, 0)$ con $x \in \mathbb{R}$: in questo senso $\mathbb{C}$ è un ampliamento di $\mathbb{R}$ (per approfondimenti sulla parte algebrica di questo argomento si veda sempre ([1], §1.3)).
Per agevolare la trattazione dei numeri complessi, identificheremo ogni coppia $(x, 0)$ con il reale x. Osserviamo allora che, per $(x, y) \in \mathbb{C}$

$$(x, y) = (x, 0) + (0, y) = (x, 0) + (y, 0)(0,1) = x + y(0,1) = x + iy,$$

dove il numero complesso $i = (0,1)$ è detto unità immaginaria. Questa rappresentazione è chiamata "forma algebrica dei numeri complessi". Ricordiamo che in $z = x + iy \in \mathbb{C}$ si ha $x, y \in \mathbb{R}$, inoltre:

- x è la parte reale di z e si indica con $Re(z)$ (o $Re\, z$, scrittura che a volte crea confusione);
- y è la parte immaginaria di z e si indica con $Im(z)$ (o $Im\, z$).

Risulta $i^2 = (0,1)(0,1) = (-1,0) = -1$ da cui $i = \sqrt{-1}$.
Il vantaggio della forma algebrica è che con i numeri complessi espressi in questo modo, si lavora con le regole usuali del calcolo polinomiale, sempre tenendo conto che $i^2 = -1$.
In questo modo è facile individuare l'inverso di un complesso non nullo $x + iy$. Dunque uno almeno tra x e y è diverso da 0, e quindi il numero reale $x^2 + y^2$ è diverso da 0, anzi positivo. Moltiplicando $x + iy$ per quello che tra poco chiameremo il suo coniugato, cioè $x - iy$, otteniamo proprio

$$(x + iy) \cdot (x - iy) = x^2 + y^2.$$

Dividendo per $x^2 + y^2$ ricaviamo

$$(x + iy) \cdot (x - iy) \cdot (x^2 + y^2)^{-1} = 1,$$

che identifica l'inverso di $x + iy$ in

$$\frac{x}{x^2 + y^2} - i\frac{y}{x^2 + y^2},$$

dove il simbolo di frazione indica il quoziente nel campo reale. L'inverso di (x, y) è $(x, y)^{-1}$ e potremo indicarlo, anche in campo complesso, con la seguente scrittura

$$(x, y)^{-1} = \frac{1}{(x, y)}.$$

In questo modo, si può estendere il discorso alla ricerca del quoziente tra due numeri complessi.

3.1.2 Complessi coniugati e modulo

Definiamo, ora, il complesso coniugato di z come il numero $\bar{z} = x - iy \in \mathbb{C}$.
Valgono le seguenti proprietà più o meno immediate.

Ci riferiamo a complessi z, w arbitrari secondo la notazione appena fissata. Si ha dunque:

- $\bar{\bar{z}} = z$;
- $z = \bar{z} \Leftrightarrow z \in \mathbb{R}$ (poiché $Im(z) = 0$);
- $z = -\bar{z} \Longrightarrow z = iy$ (poiché $Re(z) = 0$);
- $z + \bar{z} = 2 \cdot Re(z)$;
- $z - \bar{z} = 2i \cdot Im(z)$;
- $\overline{z + w} = \bar{z} + \bar{w}$,;
- $\overline{zw} = \bar{z} \cdot \bar{w}$;
- l'inverso del coniugato è uguale al coniugato dell'inverso (ci riferiamo, ovviamente, a complessi non nulli);
- per $z = x + iy$ e $\bar{z} = x - iy$, $z \cdot \bar{z} = (x + iy)(z - iy) = x^2 + y^2 = Re(z)^2 + Im(z)^2 \in \mathbb{R}$ (si noti $x^2 + y^2 \geq 0$).

Dato $z = x + iy \in \mathbb{C}$ definiamo "modulo di z" il numero reale non negativo

$$|z| = \sqrt{z \cdot \bar{z}} = \sqrt{x^2 + y^2} = \sqrt{Re(z)^2 + Im(z)^2}.$$

Il modulo gode delle seguenti proprietà; dati due complessi arbitrari z, w:

- $|z| \geq 0$ e in particolare $|z| = 0 \Leftrightarrow z = 0$;
- $|zw| = |z||w|$;
- $|Re(z)|, |Im(z)| \leq |z|$, cioè $-|z| \leq |Re(z)|, |Im(z)| \leq +|z|$;
- $|z + w| \leq |z| + |w|$ (la disuguaglianza triangolare del modulo);
- in $\mathbb{R}$ il modulo coincide con il valore assoluto;
- $|i| = |0 + i| = \sqrt{(0 + i)(0 - i)} = \sqrt{-i^2} = 1$;
- Per un qualsiasi $z = x + iy, |z| = |x + iy| \leq |x| + |iy| = |x| + |i||y| = |x| + |y|$, cioè $|Re(z)|, |Im(z)| \leq |z| \leq |Re(z)| + |Im(z)|$;
- $|z + w|^2 + |z - w|^2 = 2(|z|^2 + |w|^2)$, l'identità del Parallelogramma.

3.1.3 Rappresentazione geometrica dei numeri complessi

Fissiamo un sistema di riferimento cartesiano ortogonale oxy: ad ogni complesso $z = x + iy$ con $x, y \in \mathbb{R}$ si può associare il punto di coordinate (x, y).

Questa semplice idea fu formulata, per la prima volta, dal matematico tedesco Gauss: il piano cartesiano così ottenuto – nel quale i valori delle ascisse sono quelli della parte reale e l'ordinata rappresenta la parte immaginaria – viene chiamato anche piano di Gauss.

Ricordiamo che $|z|$ rappresenta la distanza del punto (x, y) dall'origine. Si vede molto chiaramente che $\bar{z}$ è il simmetrico di z rispetto all'asse reale ([7], §11.V).

Si rappresenta, quindi, il numero $z = x + iy$ come il vettore uscente dall'origine e con secondo estremo il punto (x, y).

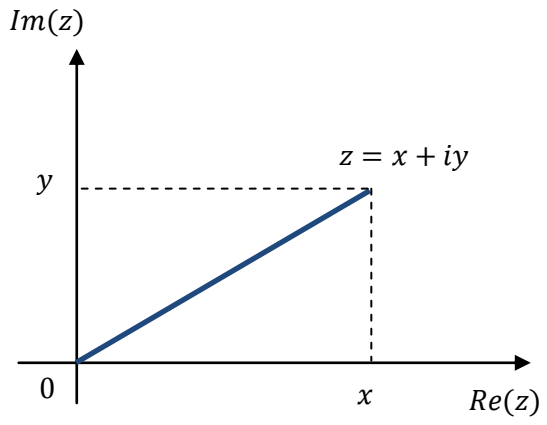

Figura 3.1. Individuazione di un valore sul piano complesso.

In questo modo la somma tra complessi corrisponde proprio alla somma vettoriale tra i rispettivi vettori e vale lo stesso per la differenza. Questa rappresentazione viene chiamata rappresentazione geometrica dei numeri complessi.
Non è di poco conto notare la corrispondenza biunivoca tra i punti del piano e i valori $z \in \mathbb{C}$.
Associando un numero complesso ad un vettore nel piano di Gauss, si rende possibile rappresentare lo stesso mediante le coordinate polari, analogamente a quanto accade ai vettori nella geometria analitica o nella fisica.
Si pone

$$\begin{cases} x = r\cos(\theta) \\ y = r\sin(\theta) \end{cases}$$

da cui $z = x + iy = r(\cos(\theta) + i\sin(\theta))$.
Questa è chiamata forma trigonometrica dei numeri complessi.

- $r = |z|$ è proprio il modulo del vettore, univocamente determinato da z;
- θ , calcolato con le usuali regole trigonometriche, è l'argomento di z e si indica anche con $\arg(z)$ (o $\arg z$) ed è univocamente determinato a meno di multipli interi di 2π.

Il risultato è quello rappresentato in Figura 3.2.

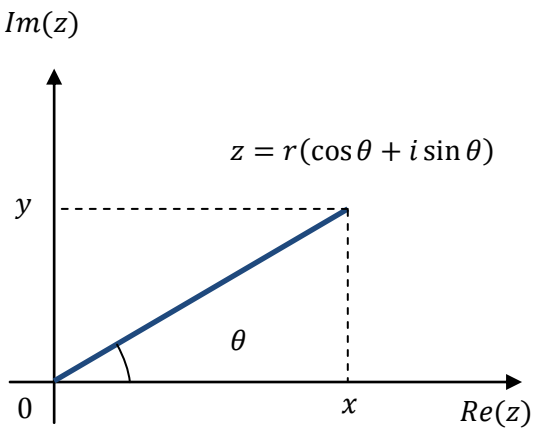

Figura 3.2. Coordinate polari per la rappresentazione geometrica dei complessi.

Possiamo osservare, inoltre, che dato $z = r(\cos(\theta) + i\sin(\theta))$, otteniamo

$$\bar{z} = r(\cos(-\theta) + i\sin(-\theta)),$$

proprio perché $\bar{z}$ è il simmetrico di z rispetto all'asse reale.

3.1.4 Prodotto e potenza n-esima di numeri complessi

Per un numero complesso $z \in \mathbb{C}$ scritto in forma algebrica, il prodotto e l'elevamento a potenza possono essere svolti in maniera agevole utilizzando le regole usuali della moltiplicazione tra binomi o dei coefficienti binomiali, ricordando sempre che $i^2 = -1$.
Questa operazione può sembrare a prima vista abbastanza immediata: possiamo considerare, ad esempio, $z = x + iy \in \mathbb{C}$ e lo si elevi al cubo. Si ottiene

$$z^3 = (x + iy)^3 = x^3 + 3x^2iy + 3xi^2y^2 + i^3y^3 = x^3 - 3xy^2 + i(3x^2y - y^3).$$

Tuttavia, al crescere dell'esponente questa procedura si rivela piuttosto difficoltosa: si pensi di voler calcolare z^{20} tanto per avere un'idea di come possano complicarsi i conti. In questo caso viene in aiuto la forma trigonometrica dei numeri complessi.
Consideriamo, a tal proposito, $z = r(\cos(\theta) + i\sin(\theta))$ e $w = s(\cos(\varphi) + i\sin(\varphi))$ due complessi in forma trigonometrica. Si ha allora

$$\begin{aligned} z \cdot w &= rs[(\cos(\theta)\cos(\varphi) - \sin(\theta)\sin(\varphi)) + i(\sin(\theta)\cos(\varphi) + \cos(\theta)\sin(\varphi))] \\ &= rs[\cos(\theta + \varphi) + i\sin(\theta + \varphi)], \end{aligned}$$

da cui si deduce che il modulo del prodotto corrisponde al prodotto dei moduli e, per quanto riguarda l'argomento, si ha $\arg(zw) = \arg(z) + \arg(w)$.

In particolare, per ogni $w = s(\cos(\varphi) + \sin(\varphi))$ complesso in forma trigonometrica, otteniamo

$$w \cdot \overline{w} = |w|^2 = s \cdot s(\cos(\varphi - \varphi) + i\sin(\varphi - \varphi)) = s^2(\cos(0) + i\sin(0)) = s^2.$$

Ne segue in particolare, per $w \neq 0$ e quindi $s \neq 0$, che l'inverso di w (ovvero il coniugato diviso il quadrato del modulo) ha forma trigonometrica

$$w^{-1} = s^{-1}[\cos(-\varphi) + i\sin(-\varphi)].$$

E' poi possibile estendere il prodotto ad un qualunque numero n di fattori: se $z_j = r_j\left(\cos(\theta_j) + i\sin(\theta_j)\right)$ con $j = 1, \ldots, n$ allora

$$\prod_{j=1}^{n} z_j = \left(\prod_{j=1}^{n} r_j\right)\left[\cos\left(\sum_{j=1}^{n} \theta_j\right) + i\sin\left(\sum_{j=1}^{n} \theta_j\right)\right]$$

Nel caso particolare $z_1 = z_2 = \ldots = z_n = z = r(\cos(\theta) + i\sin(\theta))$ si ottiene la formula per la potenza n-esima di un complesso in forma trigonometrica

$$z^n = r^n(\cos(n\theta) + i\sin(n\theta)).$$

Da notare che questa definizione continua a valere per $z = x \in \mathbb{R}$ e si estende a $z^0 = 1$. Inoltre, se n è negativo, basta applicare la formula ponendo $z^n = (1/z)^{-n}$ e la formula si estende a $\forall n \in \mathbb{Z}$.

Come già detto, l'argomento di un complesso z è unico a meno di multipli interi di 2π: in genere si fa riferimento a quello che sta in $(-\pi, \pi]$ e che è detto argomento principale di z.

3.1.5 Radici n-esime di un numero complesso

Siano $z = r(\cos(\theta) + i\sin(\theta))$ e $n \in \mathbb{N}$ fissati (con $n > 1$). L'obiettivo è cercare quei numeri complessi della forma $w = s(\cos(\varphi) + i\sin(\varphi))$ tali che $w^n = z$. Dalle definizioni precedenti, deve risultare

$$s^n(\cos(n\varphi) + i\sin(n\varphi)) = r(\cos(\theta) + i\sin(\theta)),$$

da cui $s^n = r$ e $n\varphi = \theta + 2k\pi$. Notiamo:

- $s, r \geq 0$ dunque $s = \sqrt[n]{r}$ è la radice n-esima di un reale non negativo;
- $\varphi = \frac{\theta + 2k\pi}{n}$ con $k \in \mathbb{Z}$.

In generale

$$\sqrt[n]{z} = w = \sqrt[n]{r}\left(\cos\left(\frac{\theta + 2k\pi}{n}\right) + i\sin\left(\frac{\theta + 2k\pi}{n}\right)\right), \qquad k \in \mathbb{Z}.$$

In realtà si conclude $k = 0,1,\ldots,n-1$ poiché i valori k e $k + n$ danno argomenti che differiscono di 2π e dunque corrispondono alla stessa radice ennesima di z. Si conclude che per ogni complesso $z \neq 0$ esistono esattamente n radici n-esime in $\mathbb{C}$.

Se $z = 1$ le sue radici n-esime saranno

$$\cos\frac{2k\pi}{n} + i\sin\frac{2k\pi}{n}, \quad k = 0,1,\ldots,n-1$$

Esse sono anche chiamate "radici n-esime dell'unità".

In quest'ottica le radici $n-$esime dell'unità possono essere intese come le radici del polinomio

$$x^n = 1,$$

che tra i reali ammette una soluzione (per n dispari) o due (per n pari). In $\mathbb{C}$ esistono n soluzioni distinte ovvero proprio le radici n-esime dell'unità. Vedremo in seguito che ogni polinomio di grado $n \geq 1$ a coefficienti in $\mathbb{C}$ ammette esattamente n soluzioni complesse (ognuna contata con la propria molteplicità).

3.2 FUNZIONI DI UNA VARIABILE COMPLESSA

In questa sezione parleremo di funzioni dipendenti da una variabile definita in campo complesso. Prima di questo, però, sarà necessaria una breve introduzione nella quale discuteremo sul rapporto (stretto) che esiste tra $\mathbb{C}$ e $\mathbb{R}^2$.
E' grazie ad esso che definiremo in $\mathbb{C}$ nozioni di topologia come aperti, chiusi e intorni e, per loro tramite, anche limiti e continuità per le funzioni ad una variabile complessa. Vedremo che queste ultime hanno talora proprietà radicalmente diverse da quelle delle usuali funzioni di una variabile reale. Per provarle useremo strumenti tecnici appropriati come integrazione curvilinea e teoria dei residui.

3.2.1 Topologia e successioni nel piano complesso

Dal punto di vista insiemistico, $\mathbb{C}$ è la stessa cosa di $\mathbb{R}^2$, per cui possiamo parlare di $S \subseteq \mathbb{C}$ come di $S \subseteq \mathbb{R}^2$ e viceversa ([27], §1). Inoltre il fatto che $\mathbb{C}$ è chiuso rispetto all'addizione – cioè per $z, w \in \mathbb{C}$, $z + w \in \mathbb{C}$ – e la moltiplicazione per uno *scalare* – $\alpha z \in \mathbb{C}$ con $\alpha \in \mathbb{R}, z \in \mathbb{C}$ – e la norma data dal modulo $|z| = \sqrt{x^2 + y^2}$ consentono di identificare $\mathbb{C}$ e $\mathbb{R}^2$ anche come spazi vettoriali sui reali. Grazie a questa proprietà le nozioni topologiche del piano reale si trasportano identiche al "piano" complesso.
Questa relazione tra $\mathbb{C}$ ed $\mathbb{R}^2$ era ben visibile anche dalla rappresentazione di $z \in \mathbb{C}$ come vettore nel piano di Gauss: in essa un complesso era visto come un elemento di $\mathbb{R}^2$ e questa idea è alla base delle proprietà topologiche che $\mathbb{C}$ eredita dalla sua corrispondenza con $\mathbb{R}^2$.
Definiamo $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ (detto anche piano complesso ampliato). In $\mathbb{C}$ si ammette in genere un unico tipo di ∞ al contrario, ad esempio, di $\mathbb{R}$ dove avevamo $+\infty$ e $-\infty$. Con esso operiamo in maniera seguente:

- $z + \infty = \infty + z = \infty$ per $z \in \mathbb{C}$;
- $z \cdot \infty = \infty \cdot z = \infty$ per $z \in \mathbb{C}$ e $z \neq 0$;
- $0 \cdot \infty$ e $\infty + \infty$ non sono definite e sono forme indeterminate.

Prima di passare alle funzioni, saranno utili delle definizioni di convergenza e limiti per successioni a valori complessi. Molte definizioni di base riguardo a successioni in campo complesso sono analoghe a quanto visto nel caso reale; la differenza sta soltanto nella forma.

Quando si parla di convergenza, ad esempio, il limite è, generalmente, un valore complesso quindi deve esserci convergenza nella parte reale e in quella immaginaria contemporaneamente.
Diremo che una successione $(z_n)_{n\in\mathbb{N}}$ è una funzione che ad ogni naturale (o anche ad ogni intero positivo) n associa un numero complesso z_n. In altre parole, per ogni intero $n = 0,1,2,3, \ldots$ assegniamo un valore complesso $z_0, z_1, z_2, z_3, \ldots$ ([29], §6.1).
Il numero $L \in \mathbb{C}$ è limite della sequenza $(z_n)_{n\in\mathbb{N}}$ se, $\forall\varepsilon > 0, \exists N \in \mathbb{N}$ tale che $|z_n - L| < \varepsilon$, $\forall n > N$. Se L è il limite di z_n diremo che z_n converge ad L, in altre parole

$$\lim_{n\to\infty} z_n = L,$$

o, in modo analogo

$$\lim_{n\to\infty} |z_n - L| = 0.$$

Se indichiamo $L = x + iy$ e $z_n = x_n + iy_n$, la definizione di convergenza appena enunciata si traduce in modo naturale nella seguente ([2], §10.2)

$$\lim_{n\to\infty} z_n = L \quad \Leftrightarrow \quad \lim_{n\to\infty} x_n = x, \qquad \lim_{n\to\infty} y_n = y.$$

che discende dal principio di uguaglianza dei numeri complessi. Ovviamente, ciò non accade, la successione z_n è divergente. Questa nozione di convergenza non è nuova, proprio perché la norma in $\mathbb{C}$ coincide con la distanza euclidea in $\mathbb{R}^2$.
Una successione $(z_n)_{n\in\mathbb{N}}$ in $\mathbb{C}$ è di Cauchy se $|z_n - z_m| \to 0$ per $n, m \to \infty$. In maniera equivalente, dato $\varepsilon > 0, \exists N \in \mathbb{N}$ tale che $|z_n - z_m| < \varepsilon$ per $n, m > N$. Possiamo concludere che $(z_n)_{n\in\mathbb{N}}$ in $\mathbb{C}$ è di Cauchy se lo sono anche le due successioni determinate dalla parte reale e da quella immaginaria.
Analogamente a $\mathbb{R}^2$, anche $\mathbb{C}$ è completo (§1.1.3).
Se $r > 0$ e z_0 è un numero complesso

$$D(z_0, r) = \{z \in \mathbb{C}: |z - z_0| < r\},$$

è il disco aperto – o l'intorno circolare – di centro z_0 è raggio r. Come in $\mathbb{R}^2$ esso rappresenta il luogo geometrico dei punti che hanno una distanza inferiore a r dal centro z_0. Analogamente, $\overline{D(z_0, r)} = \{z \in \mathbb{C}: |z - z_0| \leq r\}$ è il disco chiuso di centro z_0 e raggio r.
Di particolare interesse è il disco unitario per convenzione indicato con D,

$$D = \{z \in \mathbb{C}: |z| < 1\}.$$

Esso non è altro che l'interno del disco aperto di centro l'origine e raggio 1.
Valgono tutte le definizioni dette per $\mathbb{R}^2$ come intorni, aperti, chiusi, punti di accumulazione...

3.2.2 Funzioni, limiti e continuità

Le funzioni di una variabile complessa sono definite allo stesso modo delle funzioni di variabile reale ([16], §1.11).
Con abuso di notazione, possiamo indicare una funzione $f: G \subseteq \mathbb{C} \to \mathbb{C}$ tanto come $f(z)$ quanto come $f(x, y)$ assumendo $z = x + iy$. Si avrà $f(z) = f(x, y) = u(x, y) + iv(x, y)$ in cui $u(x, y)$ e $v(x, y)$ sono funzioni di due variabili reali che rappresentano, rispettivamente, $Re\big(f(x, y)\big)$ e $Im\big(f(x, y)\big)$.

Da notare, però, che questa scrittura è per lo più teorica poiché la separazione della parte reale e di quella immaginaria è una procedura generalmente complicata da operare in maniera globale.
Per una funzione semplice, quale $f(z) = z^2$, non è difficile operare in tal senso

$$z^2 = (x + iy)^2 = x^2 - y^2 + 2ixy = (x^2 - y^2) + i(2xy).$$

Tuttavia, occorre complicare anche solo un po' il problema originario per trovare più difficoltà, come ad esempio $f(z) = z^{10}$.
Se z_0 è un punto di accumulazione di G, dire che $\lim_{z \to z_0} f(z) = l \in \mathbb{C}$ significa che per ogni $\varepsilon > 0$, esiste $\delta > 0$ tale che $f(z) \in D(l, \varepsilon)$, per ogni $z \in G$ tale che $z \in D(z_0, \delta) \backslash \{z_0\}$.
In altre parole, per ogni $\varepsilon > 0$, esiste $\delta > 0$ tale che $|f(z) - l| < \varepsilon$, per ogni $z \in G \backslash \{z_0\}$ tale che $|z - z_0| < \delta$.
Diremo che $f: G \subseteq \mathbb{C} \to \mathbb{C}$ è continua in $z_0 \in G$ se per ogni $\varepsilon > 0$, esiste $\delta > 0$ tale che

$$|f(z) - f(z_0)| < \varepsilon, \qquad \text{per ogni } z \in G \text{ con } |z - z_0| < \delta.$$

Osserviamo che la definizione resta valida sempre, anche se z_0 è un punto isolato di G. Se, però, z_0 è di accumulazione, f è continua in $z_0 \Leftrightarrow \lim_{z \to z_0} f(z) = f(z_0)$.
Tutte le proprietà delle funzioni continue viste per le funzioni di variabili reali continuano a valere: somma, prodotto, composizione... di funzioni continue restano tali. Inoltre f è continua in $z_0 \Leftrightarrow Re(f)$ è $Im(f)$ sono entrambe continue in $Re(z_0)$ e $Im(z_0)$.
Con la disuguaglianza triangolare si può dimostrare che se f è continua, allora anche $|f|$ lo è ([24], §2.1), anche se non vale il viceversa.
Diremo, inoltre, che f ha un massimo in $z_0 \in G$ se $|f(z)| \leq |f(z_0)|$, per ogni $z \in G$. Invertendo la disuguaglianza abbiamo la definizione di minimo.

3.2.3 Derivabilità in senso complesso

Sia $A \subseteq \mathbb{C}$ aperto e $z_0 \in A$. La funzione $f: A \to \mathbb{C}$ si dice derivabile in senso complesso – o anche olomorfa – nel punto z_0 se esiste, finito, il limite del rapporto incrementale

$$\lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0} = \lim_{h \to 0} \frac{f(z_0 + h) - f(z_0)}{h}, \qquad h \in \mathbb{C}.$$

Se f è derivabile in senso complesso $\forall z_0 \in A$ diremo che f è olomorfa in A. Analogamente alle funzioni di una variabile reale, indicheremo con f' la derivata di f (in A).

Teorema ([24] §2.2)
Siano $f, g: A \to \mathbb{C}$ olomorfe con $A \subseteq \mathbb{C}$ aperto, allora

- $f + g$ è olomorfa in A e $(f + g)' = f' + g'$;
- fg è olomorfa in A e $(fg)' = f'g + fg'$;
- $\frac{f}{g}$ è olomorfa nei punti dove $g \neq 0$ e $\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$.

Inoltre la composizione di funzioni olomorfe è essa stessa olomorfa e vale $\left(f\big(g(z)\big)\right)' = f'\big(g(z)\big)g'(z)$ per ogni z interessato.

Valgono dunque le usuali regole di derivazione delle funzioni di variabile reale.
Prendendo la definizione di derivata una conseguenza immediata è che se f è derivabile in senso complesso è anche continua.

Equazioni di Cauchy-Riemann
Consideriamo una funzione $f: A \subseteq \mathbb{C} \to \mathbb{C}$ e assumiamo che di poter separare le variabili nella loro parte reale e immaginaria.

$$z = x + iy, \qquad f(z) = u(x,y) + iv(x,y),$$

Allora f è derivabile in senso complesso in $z_0 = (x_0, y_0)$ se e solo se le derivate parziali di u e v come funzioni di variabili reali x, y verificano le seguenti condizioni:

$$\frac{\partial u}{\partial x}(x_0,y_0) = \frac{\partial v}{\partial y}(x_0,y_0) \quad \frac{\partial u}{\partial y}(x_0,y_0) = -\frac{\partial v}{\partial x}(x_0,y_0).$$

Queste sono dette equazioni – o condizioni – di Cauchy-Riemann (C-R). Vale anche il contrario, cioè se u e v soddisfano le condizioni C-R, allora $f = u + iv$ è derivabile in senso complesso.
Per compattare la notazione, indicheremo anche $\frac{\partial u}{\partial x} = u_x, \frac{\partial u}{\partial y} = u_y, \frac{\partial v}{\partial x} = v_x, \frac{\partial v}{\partial y} = v_y$ in analogia a quanto detto nelle derivate parziali di funzioni in due variabili.

3.2.4 Successioni e serie in campo complesso

Anche per le serie in campo complesso molte definizioni sono analoghe a quanto visto nel caso reale. Tutti i risultati, inoltre, se ristretti al caso reale danno proprio definizioni e teoremi già visti su serie e successioni di valori reali.
Cominciamo con il considerare somme di un numero finito arbitrario di numeri complessi

$$\sum_{i=0}^{n} z_i = z_0 + z_1 + \ldots + z_n,$$

ma poi, come nel caso reale, ci interessiamo a serie infinite che – anche in questo caso – si possono definire come limite di una serie finita

$$\sum_{n=0}^{\infty} z_n = \lim_{m \to +\infty} \sum_{n=0}^{m} z_n.$$

Come per le serie a valori reali, anche nel caso di numeri complessi ad ogni serie si può associare una successione, cioè quella delle somme parziali

$$S_m = \sum_{n=0}^{m} z_n = z_0 + z_1 + \ldots + z_n.$$

In questo caso, la formula precedente diventa

$$\sum_{n=0}^{\infty} z_n = \lim_{m \to \infty} \sum_{n=0}^{m} z_n = \lim_{m \to \infty} S_m.$$

Diremo, dunque, che è convergente se la successione delle somme parziali è anch'essa convergente. Tuttavia questa è una definizione piuttosto teorica anche se però valgono molti

risultati e definizioni simili a quelli già discussi riguardo a serie a valori reali (convergenza assoluta, totale,...).
Diremo, ad esempio, che se $\sum_{n=0}^{\infty}|z_n|$ converge – indicandolo anche $\sum_{n=0}^{\infty}|z_n| < \infty$ – allora la serie $\sum_{n=0}^{\infty} z_n$ converge assolutamente.

Teorema
Condizione necessaria per la convergenza della serie $\sum_{n=0}^{\infty} z_n$ è che $z_n \to 0$.

Questo risultato, però, non ha molta utilità pratica proprio perché fornisce una condizione necessaria ma non sufficiente. Possiamo vederlo sotto un'altra ottica, semplicemente cambiando punto di vista. La logica insegna che $A \to B$ è equivalente a $\bar{B} \to \bar{A}$: se $z_n \nrightarrow 0$, allora $\sum_{n=0}^{\infty} z_n$ è divergente.
Diremo che $\sum_{n=0}^{\infty} z_n$ è assolutamente convergente se la serie $\sum_{n=0}^{\infty}|z_k|$ converge.
Per la convergenza valgono risultati analoghi al caso reale ([29], §6.1).

Teorema
Se $\sum_{n=0}^{\infty}|z_n|$ converge, allora anche $\sum_{n=0}^{\infty} z_n$ è convergente. In altre parole la convergenza assoluta implica quella semplice.

Teorema
Sia $\sum_{n=0}^{\infty} z_n$ una serie a termini complessi non nulli tali che

$$\lim_{n\to\infty}\left|\frac{z_{n+1}}{z_n}\right| = L.$$

Sotto questa ipotesi:

(i) se $L < 1$ allora la serie converge assolutamente
(ii) se $L > 1$ o $L = \infty$ la serie diverge
(iii) se $L = 1$ non possiamo stabilire il carattere della serie.

Teorema
Sia $\sum z_n$ una serie a termini complessi tali che

$$\lim_{\mathrm{n}\to\infty} \sqrt[n]{|z_n|} = L.$$

Sotto questa ipotesi:

(i) se $L < 1$ la serie converge assolutamente
(ii) se $L > 1$ o $L = \infty$ la serie diverge
(iii) se $L = 1$ non possiamo stabilire il carattere della serie.

3.2.5 Serie di potenze

Supponiamo di avere (a_n) una successione di numeri complessi e sia $z_0 \in \mathbb{C}$ un punto fissato. La serie

$$\sum_{n=0}^{\infty} a_n (z - z_0)^n$$

si dice serie di potenze centrata in z_0. Gli a_n sono i coefficienti della serie e con abuso di notazione intendiamo $(z - z_0)^0 = 1$ anche quando $z = z_0$ ([27], §4).
Tramite la traslazione $z \longmapsto z - z_0$, una serie di potenze di centro arbitrario z_0 si può sempre ricondurre alla serie $\sum_{n=0}^{\infty} a_n z^n$ centrata in $z_0 = 0$. In questo modo possiamo vedere che non è restrittivo lavorare con serie di quella forma proprio perché ci si può ricondurre ad essa tramite la traslazione.
Nei teoremi seguenti e nei risultati seguenti, lavoreremo sempre con serie di questa forma.

Lemma di Abel
Se $z_0 \neq 0$ e $\sum_{n=0}^{\infty} a_n z_0^n$ è convergente, allora la serie $\sum_{n=0}^{\infty} a_n z^n$ converge assolutamente (quindi totalmente) $\forall z$ con $|z| < |z_0|$. In altre parole se converge in z_0 converge anche nel disco $D(0, z_0)$. Se, invece, $\sum_{n=0}^{\infty} a_n z_0^n$ non converge, allora $\sum_{n=0}^{\infty} a_n z^n$ non converge $\forall z$ con $|z| > |z_0|$.

Il lemma di Abel è un risultato intermedio che serve per dimostrare il seguente teorema.

Teorema di convergenza delle serie di potenze
Data la serie di potenze $\sum_{n=0}^{\infty} a_n z^n$, ci sono 3 casi possibili:
- la serie converge solo in $z = 0$;
- la serie converge su tutto $\mathbb{C}$;
- $\exists R > 0$ tale che la serie converge $\forall z \in \mathbb{C}$ con $|z| < R$ e non converge per $|z| > R$.

Possiamo notare che il primo ed il secondo risultato si possono vedere come casi particolari del terzo nei quali $R = 0$ e $R = \infty$ rispettivamente. R si chiama raggio di convergenza della serie e $D(0, R)$ disco di convergenza.

Teorema
$\forall 0 < r < R$ la serie $\sum_{n=0}^{\infty} a_n z^n$ converge uniformemente nel disco chiuso $|z| < r$. La somma di questa serie è una funzione continua in $D(0, R)$.

Criterio di Hadamard
Siano $\sum_{n=0}^{\infty} a_n z^n$ e $l = \lim_{n\to\infty} \sqrt[n]{|a_n|} \in \mathbb{R}$. Allora:
- se $l = 0$, la serie $\sum_{n=0}^{\infty} a_n z^n$ converge $\forall z \in \mathbb{C}$;
- se $l = +\infty$, la serie $\sum_{n=0}^{\infty} a_n z^n$ converge solo in $z = 0$;
- se $0 < l < +\infty$, la serie $\sum_{n=0}^{\infty} a_n z^n$ converge per $|z| > \frac{1}{l}$.

Anche qui si può vedere che i primi due casi si possono rapportare al terzo come situazioni particolari di quest'ultimo.

Teorema ([20] §10.6)
La somma della serie $\sum_{n=0}^{\infty} a_n z^n$ è derivabile in $D(0, R)$ e risulta $f'(z) = \sum_{n=0}^{\infty} n a_n z^{n-1}$. Da questo segue che f è olomorfa e – iterando il procedimento – che è infinitamente derivabile.

Consideriamo un aperto $A \subseteq \mathbb{C}$. Una funzione $f: A \to \mathbb{C}$ è detta analitica se è sviluppabile in serie di potenze in un intorno di ogni punto di A. In altre parole questo equivale a richiedere che ad ogni punto $z_0 \in A$ si possano associare una successione di coefficienti $a_n = a_n(z_0)$ ed un raggio $r = r(z_0) > 0$ tali che $f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$ per $|z - z_0| < r$.
Una funzione analitica in tutto $\mathbb{C}$ è detta intera ([27], §6).

Teorema
Sia $\Omega \subseteq \mathbb{C}$ aperto e f olomorfa in Ω. Allora f è sviluppabile in serie di potenze in Ω, cioè $\forall z_0 \in \Omega$, $\exists R > 0$ ed esiste una serie di potenze $\sum_{n=0}^{\infty} a_n (z - z_0)^n$ centrata in z_0 tali che

$$D(z_0, R) \subseteq \Omega, \quad f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n \quad \forall z \in D(z_0, R).$$

Questo risultato è molto interessante e ci mostra che, sebbene ci siano non poche similitudini nella trattazione delle funzioni di una variabile complessa rispetto a quelle di una variabile reale, ci sono anche parecchie differenze.
Focalizziamoci in un particolare sulla seguente affermazione fondamentale.
"Una funzione è detta analitica se è sviluppabile in serie di potenze (definizione valida anche nel caso reale, si pensi alla serie di Taylor)".
Il teorema precedente e un risultato che vedremo in seguito ci consentiranno di affermare che nel campo complesso i termini "olomorfo" e "analitico" sono sinonimi mentre invece non troviamo un analogo riscontro, ad esempio, per funzioni di variabile reale.

3.2.6 Principio di identità per le funzioni olomorfe

Siano f e g due funzioni olomorfe in $\Omega \subseteq \mathbb{C}$ aperto. Se $\{z \in \Omega: f(z) = g(z)\}$ ha punti di accumulazione in Ω allora f è identicamente uguale a g in Ω.
Questo risultato viene chiamato anche principio del prolungamento analitico. Una conseguenza di questo principio è l'unicità del prolungamento analitico (quando esso esiste) di una funzione f di una variabile reale $x \in I$ ad un aperto connesso $\Omega \subseteq \mathbb{C}$: se I è un intervallo non degenere contenuto nell'intersezione di Ω con l'asse reale e inoltre si trova una funzione $f(z)$ analitica che coincide con f per $z = x \in I$, allora non ci può essere nessun'altra funzione analitica di z con la stessa proprietà.
In generale, la questione del prolungamento analitico è molto più ampia. Se, infatti, $f_1(z)$ e $f_2(z)$ sono due funzioni analitiche rispettivamente su $D_1, D_2 \subseteq \mathbb{C}$ con $D_1 \cap D_2 \neq \emptyset$ e $f_1 = f_2$ su $D_1 \cap D_2$, allora vale $f_1 = f_2$ su $D_1 \cup D_2$ e, dunque, f_1 (o f_2 a seconda della funzione di partenza) è l'unico prolungamento analitico di f_2 (o f_1).
A quel punto, con abuso di notazione, possiamo identificare direttamente f_1 con f_2 (o viceversa).
Vedremo che questa proprietà sarà molto importante per la ζ di Riemann: essa è una funzione analitica definita nella zona del piano complesso tale che $Re(z) > 1$, cioè nel semipiano complesso $Re(s) > 1$. Se, dunque, esiste una funzione $f(z)$ analitica definita in tutto $\mathbb{C}$ o in

una porzione più ampia di $\mathbb{C}$ rispetto alla zeta e tale che $f(z) = \zeta(z)$ per $Re(z) > 1$, allora $f(z)$ è l'unico prolungamento analitico della ζ.

3.2.7 Esponenziale e funzioni trigonometriche

Definiamo la funzione esponenziale complesso nel modo seguente:

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!} \qquad \forall z \in \mathbb{C},$$

Esso gode delle seguenti proprietà.

1. e^z è olomorfa e la sua derivata è e^z, $\forall z \in \mathbb{C}$.
2. $e^z \cdot e^w = e^{z+w}$, $\forall z, w \in \mathbb{C}$.
3. $e^z \cdot e^{-z} = e^0 = 1$, $\forall z \in \mathbb{C}$.
4. $\overline{e^z} = e^{\bar{z}}$, $\forall z \in \mathbb{C}$.
5. $\left|e^{iy}\right| = 1$, $\forall y \in \mathbb{R}$.
6. $|e^z| = e^{Re(z)}$, $\forall z \in \mathbb{C}$.

Per il principio di identità delle funzioni analitiche, l'esponenziale complesso, cioè e^z con $z \in \mathbb{C}$, è l'unico prolungamento analitico della funzione e^x con $x \in \mathbb{R}$.

Definiamo, ora, le funzioni trigonometriche.

$$\cos(z) = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!}, \qquad \sin(z) = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!}, \qquad \forall z \in \mathbb{C}.$$

Esse sono le funzioni coseno e seno definite nel campo complesso; per $z = x \in \mathbb{R}$ si hanno le usuali funzioni trigonometriche a valori reali. In analogia all'esponenziale le funzioni coseno e seno complesse sono l'unico prolungamento analitico al piano complesso delle usuali funzioni coseno e seno a valori reali rispettivamente.

Vediamo di calcolare, per $z \in \mathbb{R}$, $\cos(z) + i\sin(z)$ ricordando che $i^2 = -1$.

$$\begin{aligned}
\cos(z) + i\sin(z) &= \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!} + i \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!} \\
&= \sum_{n=0}^{\infty} i^{2n} \frac{z^{2n}}{(2n)!} + i \sum_{n=0}^{\infty} i^{2n} \frac{z^{2n+1}}{(2n+1)!} = \sum_{n=0}^{\infty} \frac{(iz)^{2n}}{(2n)!} + i \sum_{n=0}^{\infty} \frac{(iz)^{2n+1}}{(2n+1)!} \\
&= \sum_{k=0}^{\infty} \frac{(iz)^k}{k!} = e^{iz}
\end{aligned}$$

Questo è un risultato fondamentale che prende il nome di formula di Eulero. In particolare per $z = x + iy \in \mathbb{C}$

$$e^z = e^{x+iy} = e^x \cdot e^{iy} = e^x(\cos(y) + i\sin(y)).$$

Da tutto ciò segue che possiamo dare una nuova identità alle funzioni trigonometriche, definendole in termini dell'esponenziale complesso:

$$\begin{cases} e^{iz} = \cos(z) + i\sin(z) \\ e^{-iz} = \cos(z) - i\sin(z) \end{cases} \quad \Longrightarrow \quad \cos(z) = \frac{e^{iz} + e^{-iz}}{2}, \qquad \sin(z) = \frac{e^{iz} - e^{-iz}}{2i}.$$

Da questa nuova scrittura seguono in maniera banale le proprietà delle funzioni trigonometriche, in analogia a quanto accade nel caso di variabile reale. Per $z \in \mathbb{C}$:

$$\begin{aligned} \cos^2(z) + \sin^2(z) &= \frac{1}{4}\left(e^{2iz} + e^{-2iz} + 2\right) + \frac{1}{4i^2}\left(e^{2iz} + e^{-2iz} - 2\right) \\ &= \frac{1}{4}\left(e^{2iz} + e^{-2iz} + 2\right) - \frac{1}{4}\left(e^{2iz} + e^{-2iz} - 2\right) \\ &= \frac{1}{4}\left(e^{2iz} + e^{-2iz} + 2 - e^{2iz} - e^{-2iz} + 2\right) = \frac{1}{4}(2+2) = 1. \end{aligned}$$

3.2.8 Confronto con il caso reale e periodicità

Possiamo notare parecchie analogie con il caso reale per quanto riguarda le funzioni appena definite. Tuttavia, a esse, si contrappongono altrettante differenze. La prima riguarda proprio le funzioni trigonometriche che – rispetto al già citato caso di variabile reale – nel campo complesso risultano illimitate. A tale scopo consideriamo $z = iy$ con $y \in \mathbb{R}$, cioè un immaginario puro. Allora

$$\cos(iy) = \frac{e^{-y} + e^{y}}{2} \to +\infty, \qquad \text{per } y \to \pm\infty.$$

Il discorso, in realtà, è molto più ampio e variegato. Se $z = x + iy$ con $y \in \mathbb{R}$ costante reale fissata, $\cos(z)$ è una funzione periodica al variare di x, così come accade nel caso reale. Tuttavia, al variare della sua parte immaginaria, $\cos(z)$ assume un carattere completamente differente.

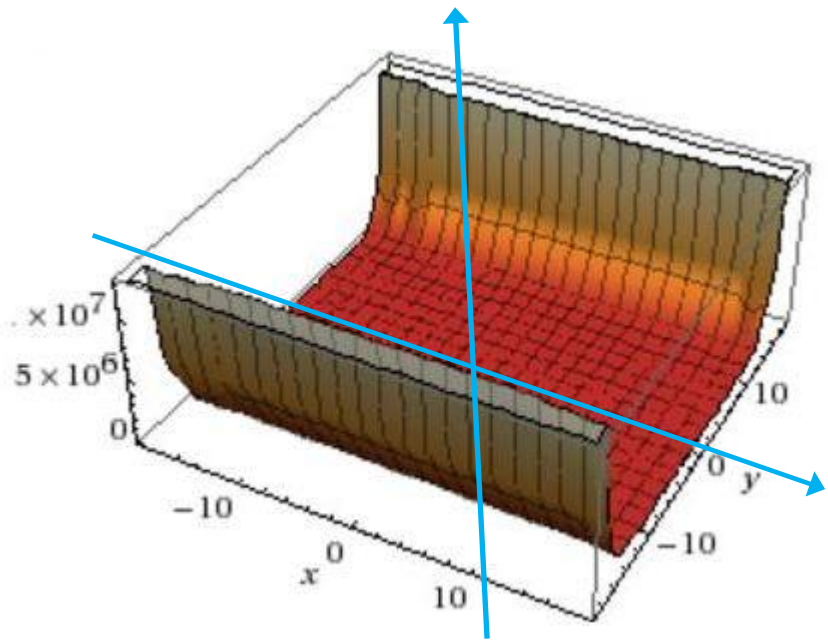

Figura 3.3. Grafico 3D di $\cos(z)$*; in azzurro sono indicati gli assi reali* x *e* y*.*

La Figura 3.3 mostra il grafico della funzione coseno di variabile complessa.
Disegnare funzioni di variabili complesse è un discorso che verrà trattato ampiamente nei capitoli successivi. Il modo più semplice e quello della Figura 3.3: al variare di $z = x + iy$ si disegna il modulo di $\cos(z)$.
Sebbene non sia uno dei procedimenti più giusti da seguire resta il più immediato: possiamo, infatti, notare come – al variare di z nella sua parte immaginaria – $|\cos(z)|$ cresca in modo *esponenziale*.
Nella Figura 3.3, in azzurro sono indicati gli usuali assi coordinati x e y con cui abbiamo a che fare nel caso di funzioni di variabili reali. Lungo di essi $|\cos(z)|$ si riconduce all'usuale $|\cos(x)|$ nel quale il modulo è inteso come valore assoluto.
La funzione coseno, dunque, è limitata se ristretta al caso di valori con parte immaginaria costante. Un discorso analogo si può fare anche con il seno complesso.
Per quanto riguarda l'esponenziale, invece, il comportamento è speculare a quello delle funzioni trigonometriche. Esso, infatti, è limitato – oltre che periodico – al variare di $z = x + iy$ con parte reale costante. Se riprendiamo la formula di Eulero, con $z \in \mathbb{C}$, $z = x + iy$

$$e^z = e^{x+iy} = e^x(\cos(y) + i\sin(y)),$$

possiamo osservare che, per $x \in \mathbb{R}$ fissato, al variare di $y \in \mathbb{R}$ la quantità contenuta all'interno delle parentesi tonde resta limitata.
Attenzione a non confondere le funzioni trigonometriche contenute nella formula di Eulero con le definizioni viste in precedenza di coseno e seno complessi: nella formula di Eulero, infatti, compaiono le usuali funzioni coseno e seno di variabile reale poiché $x, y \in \mathbb{R}$. Tuttavia si tratta solo di un cambio di punto di vista: così come per l'esponenziale, anche le funzioni trigonometriche nel campo complesso sono l'unica estensione analitica di quelle usuali di variabile reale.
Può essere una sorpresa, per il lettore abituato a lavorare con valori reali, scoprire che nel caso di variabili immaginarie il comportamento dell'esponenziale e delle funzioni trigonometriche cambi radicalmente nel piano complesso. Ad una prima analisi sembrerebbe non avere una logica: "come è possibile che l'esponenziale sia limitato al variare della parte immaginaria quando invece lungo l'asse reale non lo è?" oppure "il coseno, limitato lungo l'asse reale, come fa a non esserlo lungo quello immaginario?" possono essere domande sensate da porsi a prima vista. Il comportamento speculare di queste funzioni al variare della parte immaginaria si spiega ricordando il legame che c'è tra le stesse.
Con termini non decisamente appropriati concludiamo che definendo il coseno (o il seno) lungo valori paralleli all'asse immaginario per mezzo di un esponenziale *reale* è logico ottenere una funzione non limitata. Analogamente, tramite la formula di Eulero, si nota che lungo l'asse immaginario l'esponenziale diventa limitato, addirittura periodico.
Ricordiamo brevemente la definizione di periodicità che, in $\mathbb{C}$, è analoga a quella di funzioni di variabile reale.
Una funzione $f(z)$ è periodica se e solo se esiste $c \in \mathbb{C}$ diverso da 0 tale che $f(z + c) = f(z)$, per qualsiasi z del dominio (che dunque include anche $z + c$). Il valore di c è detto periodo di f.
Chiaramente, per ogni $k \in \mathbb{Z}$ e $z \in \mathbb{C}$, si ha che $f(z + kc) = f(z)$ poiché, a loro volta, anche i multipli interi di c sono periodi. Una funzione si dice periodica semplice se è periodica e i periodi sono i multipli interi di uno opportuno tra loro.

Considerando $c = a + ib \in \mathbb{C}$, possiamo distinguere, da un punto di vista estetico, tre tipologie di periodi per funzioni di variabile complessa:

- periodo reale, quando $c = a \in \mathbb{R}\backslash\{0\}$ (in altre parole $Im(c) = 0$);
- periodo immaginario puro, se $c = ib$ con $b \in \mathbb{R}\backslash\{0\}$ (cioè $Re(c) = 0$);
- periodo immaginario misto, se $c = a + ib \in \mathbb{C}$ con $a, b \in \mathbb{R}\backslash\{0\}$.

Per le funzioni appena introdotte, vale il seguente risultato

Teorema
La funzione e^z è periodica per immaginari puri ed i suoi periodi sono tutti e soli i numeri $2k\pi i, k \in \mathbb{Z}$. Le funzioni $\cos(z)$ e $\sin(z)$, invece, sono periodiche con periodo 2π per valori reali.

3.2.9 Osservazioni

Nell'ambito di funzioni di variabile reale, introducendo il seno ed il coseno ci si può servire – ma anche no – della circonferenza goniometrica. La circonferenza goniometrica non è altro che una particolare circonferenza centrata sull'origine ed avente raggio unitario.
La sua equazione è $x^2 + y^2 = 1$.
La peculiarità di questo luogo geometrico è che ogni punto su di esso si può individuare con una parametrizzazione particolare proprio tramite le funzioni trigonometriche usuali. Ogni punto, infatti, si può scrivere come $(\cos(t), \sin(t))$ al variare dell'angolo al centro misurato – in radianti – in senso antiorario a partire dall'asse delle x positive.
Nella totalità, questa circonferenza rappresenta il luogo dei punti aventi distanza unitaria dall'origine (distanza intesa come norma euclidea).
La Figura 3.4 serve proprio per visualizzare quanto detto.

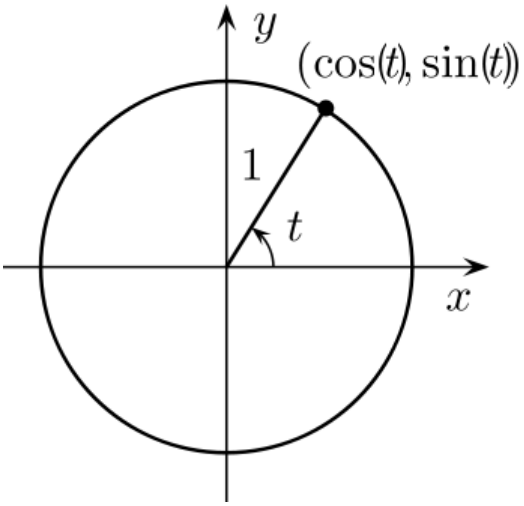

Figura 3.4. La circonferenza goniometrica.

Si può notare che l'identità trigonometrica $\cos^2(t) + \sin^2(t) = 1$ è soddisfatta: proprio per definizione, infatti, la circonferenza unitaria è l'insieme dei punti tali che $x^2 + y^2 = 1$
Nel collegare il piano complesso all'usuale piano cartesiano che rappresenta valori di $\mathbb{R}^2$ ci si può servire anche della circonferenza goniometrica notando una proprietà interessante che risulterà alla base del logaritmo definito a valori complessi. Nel piano di Gauss, infatti, possiamo individuare l'insieme dei punti aventi modulo unitario o, in analogia con $\mathbb{R}^2$, con distanza unitaria dall'origine. La distanza, in questo caso, è intesa come modulo.

Il risultato è un luogo geometrico del tutto simile alla circonferenza goniometrica: ogni punto su di esso avrà coordinate $(\cos(t), i\sin(t))$. Ritorniamo, dunque, al collegamento tra numeri complessi e piano di Gauss: ogni punto $(\cos(t), i\sin(t))$ corrisponderà ad un solo $w \in \mathbb{C}$ con norma unitaria definito come $w = \cos(t) + i\sin(t) = e^{it}$.
Concludiamo che al variare di $t \in [0,2\pi)$, il punto $w = e^{it}$ descrive, in senso antiorario, tutta la circonferenza di centro l'origine e raggio 1: per ogni $w_0 \in \mathbb{C}$ tale che $|w_0| = 1$, esiste un unico $t_0 \in [0,2\pi)$ per il quale $e^{it_0} = w_0$.

3.2.10 Funzione Logaritmo

Definiamo, ora, la funzione logaritmo: dato $w \in \mathbb{C}$, cerchiamo le soluzioni dell'equazione $e^z = w$.
Osserviamo subito che se $w = 0$ non esistono soluzioni in quanto $e^z \neq 0$, $\forall z \in \mathbb{C}$; possiamo, dunque, considerare solamente il caso $w \neq 0$.
Posto $z = x + iy$ si ha $w = e^z = e^x \cdot e^{iy}$. Segue:

- $|w| = |e^z| = e^x > 0$, si ottiene dunque $x = \log|w|$ che è il logaritmo reale del numero reale positivo $|w|$;
- $e^{iy} = w/|w|$ che è un complesso avente modulo unitario.

Per l'osservazione alla fine del paragrafo precedente esiste un unico $\theta \in [0,2\pi)$ tale che $e^{i\theta} = w/|w|$; anche tutti i valori del tipo $\theta + 2k\pi$ sono soluzione quando $k \in \mathbb{Z}$.
La conclusione è che ogni numero complesso $w \neq 0$ ha infiniti logaritmi che differiscono l'uno dall'altro per multipli interi di 2π; l'insieme di tutti questi logaritmi si indica $\log w$. In conclusione:

$$\log w = \ln|\mathrm{w}| + i(\theta + 2k\pi).$$

Dato $w \in \mathbb{C}$ non nullo, si definisce argomento del complesso w la parte immaginaria di $\log w$. Esso si indica con $\arg w$ o $\arg(w)$ ed è equivalente alla definizione data quando si parlava di rappresentazione geometrica dei numeri complessi: in questo modo, però, riusciamo a svincolarci dalla trigonometria dando una definizione che non fa uso della geometria.
Il logaritmo complesso gode delle seguenti proprietà.

- $e^{\log w} = w, \forall w \neq 0$;
- $\log e^z = z + 2k\pi i, \forall k \in \mathbb{Z}, \forall z \in \mathbb{C}$;
- $\forall w_1, w_2 \in \mathbb{C}$, $\log(w_1 w_2) = \log w_1 + \log w_2$, in particolare risulta
$$\arg(w_1) + \arg(w_2) = \arg(w_1 w_2).$$

A questo punto, $\forall w \in \mathbb{C}\backslash\{0\}$ si può scrivere nella forma $w = r \cdot e^{i\theta}$ in cui $r = |w|$ e $\theta = \arg(w)$: questa scrittura è analoga a quella vista nella rappresentazione geometrica dei numeri complessi, basta solamente ricordarsi che $e^{i\theta} = \cos\theta + i\sin\theta$.
Il logaritmo complesso, sebbene definito per $z \in \mathbb{C}\backslash\{0\}$, non è derivabile per $z \in \mathbb{R}^-$. La questione può sembrare alquanto insolita, però si spiega dalla definizione stessa di logaritmo complesso: se, infatti,

$$\log w = \ln|\mathrm{w}| + i(\theta + 2k\pi).$$

Una tale definizione, inoltre, va "unita" al fatto che, per *convenzione*, l'argomento principale di un numero complesso è $(-\pi, \pi]$, come visto in precedenza.
Questo vuol dire che se l'argomento è π, c'è un "salto" da 2π a 0 nel logaritmo, proprio perché è l'estremo di validità dell'argomento (in quanto per default l'argomento è quello principale che varia, per l'appunto, da $-\pi$ a π): questo "salto" è il motivo per cui il logaritmo complesso non è derivabile per $z \in \mathbb{R}^-$. Ricordiamo, infatti, che l'argomento in sé è un angolo e un argomento di π (o di $-\pi$) corrisponde ad un numero complesso che giace sull'asse reale negativo.
Con abuso di linguaggio, inoltre, possiamo dire che il logaritmo complesso è una funzione a più valori (in questo caso infiniti); fissando un preciso valore di $\theta = \arg w$ resta determinato un unico valore di $\log w$. Otteniamo quello che viene definito un ramo regolare: in presenza di una funzione $F(z)$ a più valori, $f(z)$ è un suo ramo regolare (in un dominio D) se f è continua in D e se $\forall z \in D$, $f(z)$ coincide con uno dei valori di $F(z)$.
In un dominio D, un ramo regolare non è altro che una restrizione di una funzione a più valori; restrizione operata in modo che la mappa ottenuta assuma un unico valore al variare di $z \in D$.

3.2.11 Potenze con esponente complesso

$\forall z \in \mathbb{C}\backslash\{0\}$ e $w \in \mathbb{C}$ definiamo

$$z^w = e^{w \log z} = e^{w\{\ln|z| + i \arg z\}} = e^{w\{\ln|z| + i(\theta + 2k\pi)\}}, \qquad k \in \mathbb{Z},$$

nella quale θ è uno specifico argomento scelto. In generale z^w è una funzione ad infiniti valori ed ogni ramo regolare del logaritmo determina un ramo regolare di z^w. Gli infiniti valori di z^w differiscono l'uno dall'altro per il fattore $e^{2k\pi i w}$.

- Se $w = n \in \mathbb{Z}$, allora $e^{2kn\pi i} = 1$, $\forall k$; z^n è una funzione ad un solo valore e si ottiene la formula vista inizialmente nel caso di esponente intero.
- Se $w = m/n$, con $m, n \in \mathbb{Z}$: $(m, n) = 1$ allora $e^{2k\pi \frac{m}{n} i}$ ha n valori distinti che si ottengono per $k = 0, \dots, n-1$; $z^{\frac{m}{n}}$ assume un numero finito di valori.

A questo punto un'osservazione fondamentale: $\forall w \in \mathbb{C}$ non nullo, $\exists z, v \in \mathbb{C}$ tali che $z^v = w$. Fissato z troviamo un esponente v tale per cui $z^v = w$ oppure fissato v possiamo trovare z tale che $z^v = w$. Questo vuol dire che, al contrario di ciò che accade nel caso di valori reali, in campo complesso $z^v = w$ ha sempre almeno una soluzione per qualsiasi scelta di z e w non nulli.
In quest'ottica, risulterà sensato ricercare le radici di equazioni del tipo $\sum_{k=0}^{n} z^k = 0$ (sempre con $k \neq n$). Un discorso del genere non aveva proprio senso nel caso di variabile reale dal momento che una somma di esponenziali non poteva essere nulla: in $\mathbb{R}$, infatti, $a^x > 0$ per ogni scelta di $a, x \in \mathbb{R}$ mentre abbiamo appena visto che il campo complesso oltrepassa questa limitazione offrendo orizzonti pressoché sconfinati.
A tale proposito offriamo un paio di esempi: calcoliamo 2^i. Otteniamo:

$$2^i = e^{i \log 2}.$$

In questo caso il logaritmo che compare nell'esponenziale è quello usuale di variabile reale; da notare che $i \log 2$ è un immaginario puro.
Ricordiamo la formula di Eulero, per $z \in \mathbb{C}$ $z = x + iy$ con $x, y \in \mathbb{R}$ si ha

$$e^z = e^x(\cos(y) + i\sin(y)),$$

Allora:

$$e^{i \log 2} = \cos(\log 2) + i\sin(\log 2).$$

L'esempio classico, invece, è il seguente: calcoliamo $e^{i\pi}$.

$$e^{i\pi} = \cos(\pi) + i\sin(\pi) = -1.$$

L'identità $e^{i\pi} + 1 = 0$ prende il nome di Identità di Eulero e serve proprio a testimoniare che, diversamente dall'usuale comportamento delle funzioni di variabile reali, in campo complesso l'esponenziale può anche assumere valori reali negativi.

3.3 INTEGRAZIONE COMPLESSA

In questa sottosezione chiariremo come integrare una funzione di una variabile complessa lungo una curva. Seguiranno, nella prossima sottosezione, molti risultati che mostreranno come verificare il carattere di una funzione (per esempio se è olomorfa o se è integrabile) in base a risultati dovuti all'integrazione di questa lungo delle curve chiuse.

3.3.1 Curve in $\mathbb{C}$

Una curva (piana) γ in $\mathbb{C}$ è una funzione continua definita su un intervallo $[a, b] \subseteq \mathbb{R}$, in modo simile a quanto detto per curve in $\mathbb{R}^2$. A seconda di quale notazione definiamo, la si può indicare come

$$[a, b] \ni t \longmapsto \gamma(t) = x(t) + iy(t)$$

oppure

$$[a, b] \ni t \longmapsto \gamma(t) = \big(x(t), y(t)\big).$$

Delle due scritture, la seconda è presa dall'usuale interpretazione di curva parametrica in $\mathbb{R}^2$ ([27], §8). Essa viene detta differenziabile – o liscia ([24] §1.3) – se $\gamma'(t) = x'(t) + iy'(t)$ esiste in tutti i punti ed è una funzione continua $\forall t \in [a, b]$.
Vediamo qualche altra definizione riguardante le curve in $\mathbb{C}$: si può trovare riscontro con quanto detto per le curve in $\mathbb{R}^2$.

- γ si dice regolare se è differenziabile e $\gamma'(t) \neq 0, \forall t \in \mathbb{C}$.
  Per questa definizione occorre fare attenzione: con abuso di scrittura, possiamo dire che una curva parametrica non è altro che una funzione a due variabili dipendenti dal parametro t.
  Dire "$\gamma'(t) = 0$" equivale a richiedere che si annulli contemporaneamente nella sua parte reale e in quella immaginaria. Supponiamo di avere $\gamma(t) = \cos(t) + i\sin(t)$,

per $t \in [0,2\pi]$ essa non è altro che la parametrizzazione del cerchio unitario. Si può facilmente osservare che $\gamma'(t) = -\sin(t) + i\cos(t) \neq 0$ per ogni $t \in [0,2\pi]$ poiché $\sin(t)$ e $\cos(t)$ non si annullano mai contemporaneamente.

- γ si dice regolare a tratti se esiste una partizione di $[a,b]$ tale che γ è regolare in ciascun intervallo della partizione. Essa si può intendere come un unione di curve regolari in ciascun intervallo della partizione: $\gamma = \gamma_1 + \gamma_2 + \cdots + \gamma_n$, nel quale n è il numero di intervalli della partizione stessa.
  Una curva regolare a tratti è detta anche cammino ([27], §8). In questa tesi la chiameremo pure catena (anche se i termini, cammino e catena, sono comunemente intesi in modo sottilmente diverso – non approfondiremo qui la distinzione).
- γ si dice chiusa se $\gamma(a) = \gamma(b)$.
- γ si dice semplice se non ha auto-intersezioni, cioè se $\gamma(t_1) \neq \gamma(t_2)$, $\forall t_1, t_2 \in [a,b]$ con $t_1 \neq t_2$ (escludendo, al più, $t_1 = a$ e $t_2 = b$).
  Una curva semplice chiusa si dice curva di Jordan.
- Un ciclo è una catena nella quale le curve (regolari) che lo compongono sono chiuse.
- Due curve γ e $\tilde{\gamma}$, con $\gamma: [a,b] \to \mathbb{C}$ e $\tilde{\gamma}: [c,d] \to \mathbb{C}$ si dicono equivalenti se esiste una φ suriettiva e derivabile, $\varphi: [c,d] \to [a,b]$ t.c. $\varphi'(\tau) \neq 0$, $\forall \tau \in [c,d]$ e $\tilde{\gamma}(\tau) = \gamma(\varphi(\tau))$, $\forall \tau \in [c,d]$. Due curve equivalenti hanno la stessa immagine (o traccia) in $\mathbb{C}$. Diremo che hanno lo stesso verso se $\varphi'(\tau) > 0$, $\forall \tau$ o verso opposto se $\varphi'(\tau) < 0$, $\forall \tau$ in cui $\tau \in [c,d]$.

Uno degli esempi più semplici – e più ricorrenti – di curva in $\mathbb{C}$ è la circonferenza unitaria di centro l'origine e raggio 1 (nel piano di Gauss). Essa, scrivibile proprio come il luogo geometrico dei punti aventi centro 0 e raggio 1, quindi $|z| = 1$, è parametrizzata nel modo seguente

$$\gamma = \cos(t) + i\sin(t) = e^{it}, \qquad t \in [0,2\pi].$$

Se ne era discusso in precedenza, mostrando come per ogni punto w tale che $|w| = 1$, esistesse un valore $t_0 \in [0,2\pi]$ tale per cui $w = e^{it_0}$ (§3.2.9).

3.3.2 Integrale su una curva

Consideriamo $A \subseteq \mathbb{C}$ aperto e sia $f: A \to \mathbb{C}$ continua; se $\gamma: [a,b] \to \mathbb{C}$ è tale che $\gamma([a,b]) \subseteq A$ e γ regolare, definiamo

$$\int_\gamma f(z)dz = \int_a^b f(\gamma(t))\gamma'(t)dt$$

che è un'integrale fatto su una curva parametrica. Questo integrale non dipende dalla parametrizzazione scelta e soddisfa le usuali proprietà degli integrali (linearità rispetto alla somma algebrica, …).

Se γ è una curva lineare a tratti, allora $\gamma = \gamma_1 + \gamma_2 + \ldots + \gamma_n$ con γ_i regolari per $i = 1, \ldots, n$ e si pone:

$$\int_\gamma f(z)dz = \int_{\gamma_1} f(z)dz + \ldots + \int_{\gamma_n} f(z)dz$$

Vediamo di fare un esempio: supponiamo di voler calcolare l'integrale di $f(z) = z$ lungo la circonferenza unitaria. Proporremo due differenti modi (che risultano essere equivalenti).
Consideriamo, per $t \in [0,2\pi]$, $\gamma(t) = \cos(t) + i\sin(t)$, $\gamma'(t) = -\sin(t) + i\cos(t)$, allora

$$\begin{aligned}\int_\gamma z\,dt &= \int_0^{2\pi} (\cos(t) + i\sin(t))(-\sin(t) + i\cos(t))dt \\ &= \int_0^{2\pi} (-\cos(t)\sin(t) - i\sin^2(t) + i\cos^2(t) - \sin(t)\cos(t))dt \\ &= -\int_0^{2\pi} \cos(t)\sin(t)\,dt - i\int_0^{2\pi} (\sin^2(t) - \cos^2(t))dt \\ &- \int_0^{2\pi} (\sin(t)\cos(t))dt = -2\int_0^{2\pi} (\cos(t)\sin(t))dt - i\int_0^{2\pi} -\cos(2t)\,dt \\ &= -\int_0^{2\pi} \sin(2t)\,dt + i\int_0^{2\pi} \cos(2t)\,dt = \frac{\cos(2t)}{2}\bigg|_0^{2\pi} + i\frac{\sin(2t)}{2}\bigg|_0^{2\pi} \\ &= \frac{1}{2} - \frac{1}{2} + i(0-0) = 0.\end{aligned}$$

Ora, come parametrizzazione alternativa della circonferenza unitaria, consideriamo $\gamma(t) = e^{it}$ sempre con $t \in [0,2\pi]$. In essa $\gamma'(t) = ie^{it}$ dalle usuali regole di derivazione. Allora

$$\int_\gamma zdt = \int_0^{2\pi} \left(e^{it} \cdot \left(ie^{it}\right)\right) dt = \int_0^{2\pi} ie^{2it}dt = \frac{1}{2}e^{it}\bigg|_0^{2\pi} = \frac{1}{2}e^{2i\pi} - \frac{1}{2} = \frac{1}{2} - \frac{1}{2} = 0.$$

Questo semplice esempio vuole essere solamente indicativo per ciò che riguarda il modo di operare quando abbiamo a che fare con integrali curvilinei in campo complesso. Si vedrà che non sarà un caso il fatto che l'integrale di $f(z) = z$ (olomorfa) su una curva chiusa risulta essere nullo.
Enunciamo ora alcune proprietà dell'integrale su una curva.

- $\int_\gamma f(z)dz = \overline{\int_\gamma \overline{f(z)}dz}$;
- $\int_\gamma f(z)dx = \frac{1}{2}\int_\gamma f(z)dz + \frac{1}{2}\int_\gamma f(z)d\bar{z}$ e $\int_\gamma f(z)dy = \frac{1}{2i}\int_\gamma f(z)dz - \frac{1}{2i}\int_\gamma f(z)\,d\bar{z}$;
- $\int_\gamma |dz| = \int_a^b |\gamma'(t)|dt = l(\gamma)$ è la lunghezza di γ;
- $\int_\gamma f(z)|dz| = \int_a^b f(\gamma(t))|\gamma'(t)|dt$ è l'integrale fatto rispetto alla lunghezza d'arco;
- $\left|\int_\gamma f(z)dz\right| \leq \int_\gamma |f(z)||dz|$, in particolare se $|f(z)| \leq M$, $\left|\int_\gamma f(z)dz\right| \leq M\int_\gamma |dz| = Ml(\gamma)$.

3.3.3 L'indice di avvolgimento e le sue proprietà

Consideriamo $\gamma \subseteq \mathbb{C}$ una curva chiusa regolare a tratti e $z_0 \in \mathbb{C}$. Indichiamo con $Ind_\gamma(z_0)$ l'indice di avvolgimento di γ rispetto al punto z_0; esso è il numero *intero*

$$Ind_\gamma(z_0) = \frac{1}{2\pi i}\int_\gamma \frac{1}{z - z_0}dz.$$

Lo chiameremo anche indice di z_0 rispetto a γ: in senso geometrico, rappresenta proprio il numero di volte che la curva chiusa γ si "avvolge" intorno al punto z_0.
Ora, il concetto di "avvolgimento" di una curva $\gamma: [a, b] \to \mathbb{C}$ intorno ad un punto non è difficile da capire. Siccome la curva è chiusa, γ è tale che $\gamma(a) = \gamma(b)$. Rispondere alla domanda "quante volte γ si avvolge ad un punto z_0 interno ad essa" equivale a dire quante volte γ lo circonda completamente percorrendo la curva da $\gamma(a)$ a $\gamma(b)$. Se γ è una circonferenza, ad esempio, l'indice di avvolgimento è sempre 1 per ogni punto interno ad essa.
Vediamo qualche proprietà dell'indice.

- $Ind_\gamma(z_0) = Ind_{-\gamma}(z_0)$, in altre parole l'indice resta lo stesso anche se si cambia il verso di percorrenza della curva γ.
- $Ind_\gamma(z_0) = 0$ per ogni punto esterno a D con $\gamma \subseteq D$. Infatti, se z_0 è esterno alla curva, la funzione $\frac{1}{z-z_0}$ è olomorfa, cioè analitica, e il suo integrale è nullo (per il teorema di Cauchy che vedremo a breve).
- Consideriamo, ora, $\Omega = \mathbb{C} \backslash \gamma([a, b])$, cioè l'insieme dei complessi a cui si sottrae quello composto dai punti della curva γ chiusa regolare a tratti. Si nota subito che Ω è formato da tante componenti connesse: su ognuna di esse l'indice è costante.
- Se γ, oltre ad essere chiusa, è anche semplice allora $Ind_\gamma(z_0) = 1$ per ogni punto z_0 interno a γ. Infatti una curva semplice non ha auto-intersezioni e quindi non si avvolge più di una volta intorno ad un qualsiasi punto interno.

Sia ora Ω un aperto di $\mathbb{C}$. Diremo che un ciclo $\gamma \subseteq \Omega$ è omologo a zero modulo Ω – e lo indicheremo con $\gamma \sim 0$ (mod Ω) – se $Ind_\gamma(z_0) = 0$ per ogni $z_0 \notin \Omega$, in altre parole γ non gira intorno a nessun punto esterno ad Ω. Diremo, inoltre, che due cicli γ_1, γ_2 contenuti in Ω sono omologhi tra loro modulo Ω – cioè $\gamma_1 \sim \gamma_2$ (mod Ω) – se il ciclo $\gamma = \gamma_1 - \gamma_2$ è omologo a zero modulo Ω.

3.3.4 Risultati importanti sulla integrazione complessa

Enunciamo in questa sezione alcuni risultati importanti – senza dimostrazione – per quanto riguarda l'integrazione complessa lungo curve. Sono risultati – taluni anche affascinanti – che trovano riscontri negli integrali curvilinei di funzioni di due variabili reali: questo è anche piuttosto ovvio vista l'analogia tra $\mathbb{C}$ ed $\mathbb{R}^2$.

Teorema ([24], §1.3)
Se una funzione continua f ha una primitiva F in Ω e $\gamma \subseteq \Omega$ è una curva che inizia in w_1 e termina in w_2 con $w_1, w_2 \in \Omega$, allora

$$\int_\gamma f(z)dz = F(w_2) - F(w_1).$$

Teorema (Cauchy)
Se $A \subseteq \mathbb{C}$ è un dominio semplicemente connesso e $f: A \to \mathbb{C}$ è analitica allora $\int_\gamma f(z)dz = 0$ per ogni $\gamma \subseteq A$ curva chiusa regolare a tratti.

Teorema
Una funzione olomorfa f in un disco aperto ha una primitiva in esso.

Teorema
Sia A semplicemente connesso. Allora $f: A \to \mathbb{C}$ ammette una primitiva in A se $\int_\gamma f(z)dz = 0$ per ogni curva $\gamma \subseteq A$ chiusa regolare a tratti.

Teorema (Cauchy in forma generale)
Sia f olomorfa in un aperto Ω contenuto in $\mathbb{C}$ allora le seguenti affermazioni sono equivalenti.

1. $\int_\gamma f(z)dz = 0$, $\forall \gamma$ ciclo contenuto in Ω con $\gamma \sim 0$ (mod Ω).
2. Se $\gamma \sim 0$ (mod Ω), allora $\forall z \in \Omega \backslash \gamma$ si ha $f(z) \cdot Ind_\gamma(z) = \frac{1}{2\pi i} \int_\gamma \frac{f(w)}{w-z} dw$.
3. Se $\gamma_1 \sim \gamma_2$ (mod Ω), allora $\int_{\gamma_1} f(z)dz = \int_{\gamma_2} f(z)dz$.

I teoremi precedenti sono tutti collegati e derivano da una matrice comune. La generalizzazione di questi risultati è il seguente

Teorema ([27], §13)
$A \subseteq \mathbb{C}$ è semplicemente connesso se e solo se ogni funzione $f: A \to \mathbb{C}$ analitica ha una primitiva in A.

Teorema (Formula integrale di Cauchy nel disco)
Siano f una funzione olomorfa in un disco D e γ una curva chiusa, regolare a tratti e contenuta in D. Allora per ogni punto $z_0 \in \mathbb{C} \backslash \gamma$ si ha

$$f(z_0) \cdot Ind_\gamma(z_0) = \frac{1}{2\pi i} \int_\gamma \frac{f(z)}{z - z_0} dz,$$

dove l'integrale lo si considera percorso in senso antiorario.

Questa formula è molto importante, soprattutto nel caso in cui abbiamo a che fare con una curva semplice poiché $Ind_\gamma(z_0) = 1$. Questo implica che se f è olomorfa in un disco – caso particolare di una curva semplice – allora i valori di f all'interno sono completamente determinati da quelli della frontiera.

Teorema ([27] , §11; [16], §9.4)
Se $\Omega \subseteq \mathbb{C}$ è un aperto e f è olomorfa in Ω, allora f è analitica, cioè sviluppabile in serie di potenze in Ω. In altre parole $\forall z_0 \in \Omega$, esiste $R > 0$ ed esiste una serie di potenze centrata in z_0 tali che

$$D(z_0, R) \subseteq \Omega, \quad f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n \quad \forall z \in D(z_0, R)$$

e la serie trovata coincide con quella di Taylor di centro z_0. Inoltre, $\forall z_0 \in \Omega$, si ha

$$f^{(n)}(z_0) = \frac{n!}{2\pi i}\int_\gamma \frac{f(w)}{(w-z_0)^{n+1}}\,dw.$$

Questo teorema è la chiusura del cerchio: si era visto il viceversa in ambito delle serie di potenze ma ora sappiamo che è ambivalente. In altre parole una funzione f è olomorfa in Ω se e solo se è analitica cioè essa è la somma di una serie di potenze.
Si era già osservato che a differenza del caso reale la nozione di derivabilità complessa implica che una funzione f di variabile complessa è C^∞. Ora si va oltre, cioè se f è olomorfa, allora esiste una serie di potenze che ha come somma f. Nel caso reale questa proprietà non vale: possiamo ad esempio considerare la funzione f così definita

$$f(x) = \begin{cases} e^{-\frac{1}{x^2}}, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

che è C^∞ ma non analitica.
Inoltre, se $f: \mathbb{C} \to \mathbb{C}$ è olomorfa su tutto il piano complesso, allora lo sviluppo $f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} z^n$ centrato nell'origine vale per qualsiasi $z \in \mathbb{C}$ mentre la stessa cosa non si registra in campo reale. Per esempio, $f(x) = \frac{1}{1+x^2}$ con $x \in \mathbb{R}$ ha $f(z) = \sum_{n=0}^{\infty}(-1)^n x^{2n}$ che vale solo per $|x| < 1$.
Un fatto curioso, invece, è che se f è olomorfa in vari domini isolati, in ognuno di questi potrebbe avere uno sviluppo in serie di potenze differente.

Stima di Cauchy sulle derivate ([24], §9.2)
Sia f olomorfa in $\overline{D(z_0, R)}$ e sia $|f(z)| \leq M$, $\forall z \in \partial D(z_0, R)$ con M costante positiva. Allora

$$\left|f^{(n)}(z_0)\right| \leq \frac{M \cdot n!}{R^n}$$

Teorema (Morera)
Sia f continua in un dominio Ω. Se $\int_\gamma f(z)dz = 0$ per ogni curva γ chiusa, regolare a tratti e contenuta in Ω, allora f è olomorfa in Ω.

Teorema (media di Gauss o proprietà integrale della media) ([27], §12; [16], §9.10)
Se f è olomorfa in $\overline{D(z_0, R)}$, allora

$$f(z_0) = \int_0^{2\pi} f\left(z_0 + Re^{it}\right)dt.$$

In altre parole il valore di R nel centro è la media dei valori fatti sul bordo. Il risultato segue immediatamente dalla formula integrale di Cauchy considerando il bordo del disco come una curva chiusa semplice.

3.4 SVILUPPO DI LAURENT, ZERI E SINGOLARITA'

Vedremo, per una funzione di variabile complessa, lo sviluppo di Laurent come strumento per introdurre il concetto di singolarità e i vari tipi di singolarità per una funzione.

3.4.1 Sviluppo di Laurent

Lo sviluppo di Laurent è un passo fondamentale nella comprensione di molte proprietà particolari delle funzioni di variabile complessa. Esso reca con sé molte problematiche ed altrettanti spunti per un'analisi approfondita di tali questioni: tuttavia in questa sezione sarà trattato solamente in maniera essenziale per cogliere i tratti fondamentali dei concetti che ci interessano, le singolarità.
Sia una funzione f olomorfa in Ω un dominio di $\mathbb{C}$. Definiamo lo sviluppo di Laurent di f nel modo seguente

$$f(z) = \sum_{n=-\infty}^{n=\infty} c_n (z - z_0)^n.$$

In questa rappresentazione:

$$c_n = \frac{1}{2\pi i} \int_\gamma \frac{f(w)}{(w - z_0)^{n+1}} dw, \qquad \forall n \in \mathbb{Z}$$

nel quale γ è una qualunque circonferenza di centro z_0 e raggio r percorsa in senso antiorario (e, ovviamente, contenuta in Ω). Questo sviluppo esiste ed è unico ([16], §9.17-9.18; [27], §14).

3.4.2 Zeri di una funzione di variabile complessa

Diremo che il punto $z_0 \in A$ è uno zero della funzione $f: A \subseteq \mathbb{C} \to \mathbb{C}$ se $f(z_0) = 0$.
Diremo, inoltre, che m è l'ordine dello zero di f se $f(z_0) = g(z)(z - z_0)^m$ con g olomorfa e $g(z_0) \neq 0$: dire che z_0 è uno zero di ordine m equivale a dire che nel punto z_0 si annullano, oltre alla f, anche tutte le derivate fino all'ordine $m - 1$ in analogia al caso di funzioni di variabile reale.
Se $m = 1$ lo zero viene detto semplice.
Si mostreranno, ora, alcuni risultati riguardanti gli zeri di funzioni olomorfe.

Teorema ([20], §10.18; [24], §3.1)
Siano Ω un dominio di $\mathbb{C}$ e f una funzione olomorfa in Ω.

Se $Z(f)$ è l'insieme degli zeri di f in Ω, cioè

$$Z(f) = \{z \in \Omega: f(z) = 0\},$$

allora o $Z(f) = \Omega$ oppure $Z(f)$ non può avere punti di accumulazione.

In altre parole questo teorema ci dice che gli zeri di f sono solamente dei punti isolati altrimenti la funzione sarebbe identicamente nulla; inoltre, $\forall z \in \Omega$,e $z_0 \in Z(f)$, $\exists!\, m$ intero positivo, g olomorfa con $g(z_0) \neq 0$ tali che $f(z) = (z - z_0)^m g(z)$ in accordo alla definizione data in precedenza.

Teorema fondamentale dell'algebra
Un polinomio $p(z)$ – con $z \in \mathbb{C}$ – di grado $m \geq 1$ ha m zeri (in $\mathbb{C}$) contati con la loro molteplicità.

Teorema (Rouché)
Sia $\gamma \sim 0 \pmod{\Omega}$ un ciclo contenuto in $\Omega \subseteq \mathbb{C}$ aperto tale che $Ind_\gamma(z) = 0,1$, $\forall z \in \Omega$. Siano, inoltre, f, g funzioni olomorfe in Ω tali che

$$|f(z) - g(z)| < |f(z)| \qquad \forall z \in \gamma.$$

Allora f e g hanno lo stesso numero di zeri nel dominio $\Omega_1 = \{z \in \Omega: Ind_\gamma(z) = 1\}$.

Questo teorema è importante perché, grazie ad una semplice maggiorazione, permette di calcolare il numero di zeri all'interno di un ciclo di una funzione, magari di non facile studio.
Vediamo di fare un esempio: come ciclo consideriamo la circonferenza unitaria, cioè $|z| = 1$ – o $\gamma = e^{it}$ se si preferisce –, e come funzione $g(z) = z^7 - 5z^3 + 12$. Il dominio che ci interessa è $|z| < 1$, cioè il disco unitario che non è altro che l'interno della circonferenza unitaria.
A questo punto, riprendendo la notazione del teorema di Rouché, poniamo $g(z) = z^7 - 5z^3 + 12$ e $f(z) = 12$. Allora

$$|f(z) - g(z)| = |12 - z^7 + 5z^3 - 12| = |-z^7 + 5z^3| \leq |z|^7 + 5|z^3| = 6 \qquad \forall z \in \gamma$$

$$|f(z)| = 12 \qquad \forall z$$

La condizione del teorema di Rouché, cioè

$$|f(z) - g(z)| < |f(z)|,$$

è rispettata proprio perché $6 < 12$ quindi possiamo concludere che $f(z)$ e $g(z)$ hanno lo stesso numero di zeri all'interno del dominio considerato, cioè la circonferenza unitaria. Nel nostro caso il numero di zeri è... nessuno! Infatti $f(z) = 12 \neq 0$ essendo f una funzione costante.

3.4.3 Singolarità isolate

Indichiamo, per semplificare la notazione, con $\widetilde{D}(z_0, R)$ il disco bucato di centro z_0 e raggio R; in altre parole non è altro che $D(z_0, R) \backslash \{z_0\}$, cioè il cerchio meno il punto centrale,

$$\widetilde{D}(z_0, R) = \{z \in \mathbb{C}: 0 < |z - z_0| < R\}.$$

Sia Ω un aperto di $\mathbb{C}$; se f è olomorfa in $\Omega\backslash\{z_0\}$ diremo che z_0 è una singolarità isolata di f.
Per esempio la funzione

$$f(z) = \frac{1}{z},$$

è olomorfa in $\mathbb{C}\backslash\{0\}$ e lo zero è una sua singolarità isolata.
Per i teoremi precedenti, f è sviluppabile in serie di Laurent in $\widetilde{D}(z_0, R) \subseteq \Omega$. Sia, dunque,

$$f(z) = \sum_{n=-\infty}^{\infty} a_n (z - z_0)^n$$

lo sviluppo di Laurent di f in $\widetilde{D}$. Ci sono 3 casi possibili.

- Tutti i coefficienti a_n dello sviluppo di Laurent con n negativo sono nulli, cioè $f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$. In questo caso z_0 è una singolarità eliminabile per f: infatti la funzione la si può estendere in modo olomorfo in tutto il disco $D(z_0, R)$ definendo $f(z_0) = a_0$.
- Solo un numero finito di coefficienti a_n con n negativo è diverso da zero così che lo sviluppo è del tipo $\sum_{n=-k}^{\infty} a_n (z - z_0)^n$ con k intero positivo.
  In questo caso diremo che z_0 è un *polo* di f; se, inoltre, succede che $a_{-k} \neq 0$ e $a_{-l} = 0, \forall l \geq k + 1$ allora il polo è di ordine k. Se $k = 1$ il polo è semplice.
- Se ci sono infiniti coefficienti a_n con n negativo diversi da zero, la singolarità z_0 viene detta essenziale per f.

Vorremo richiamare il seguente risultato sulle singolarità di tipo polo per una funzione olomorfa che sarà utile nella trattazione della funzione ζ di Riemann: se $z_0 \in \mathbb{C}$ è un polo di ordine m per f allora in z_0 la funzione si può scrivere nel seguente modo

$$f(z) = \frac{g(z)}{(z - z_0)^m}, \qquad g(z) \text{ olomorfa e non possiede poli in } z_0,$$

inoltre $\lim_{z \to z_0} f(z) = \infty$.
Se abbiamo $f(z)$ e $g(z)$ funzioni intere – cioè olomorfe in tutto il dominio – allora la funzione

$$\frac{f(z)}{g(z)}$$

è una funzione che possiede singolarità isolate eliminabili oppure di tipo polo. Una funzione con queste proprietà viene detta meromorfa.

3.5 RESIDUI

In quest'ultima sottosezione, parleremo di un concetto avanzato di analisi complessa che ci mostra nuovamente come quest'ultima si discosti in maniera significativa da quella reale. Infatti si troverà un collegamento tra il calcolo di integrali su curve chiuse e i poli della funzione integranda.
Questo collegamento tra due concetti apparentemente distanti è dato proprio dai residui di una funzione meromorfa.

3.5.1 I residui e il teorema dei residui

Siano f olomorfa in $\widetilde{D} = \widetilde{D}(z_0, R)$ e $f(z) = \sum_{n=-\infty}^{\infty} a_n (z - z_0)^n$ il suo sviluppo di Laurent centrato in $z_0 \in \widetilde{D}$. Il coefficiente a_{-1} di questo sviluppo viene chiamato residuo di f in z_0 e si indica con $Res(f, z_0)$. Ricordando la formula per i coefficienti vista in precedenza, si ha

$$a_{-1} = Res(f, z_0) = \frac{1}{2\pi i} \int_\gamma f(z) dz$$

nel quale γ è la circonferenza di centro z_0 e raggio $r \in (0, R)$ percorsa in senso antiorario.

Teorema dei residui
Sia f olomorfa in un aperto Ω escluse al più delle singolarità isolate z_j con $j \in \{1, \dots, n\}$. Sia anche γ una curva chiusa semplice, regolare a tratti contenuta in Ω che non passa per nessuno dei punti z_j.
Allora

$$\int_\gamma f(z) dz = 2\pi i \sum_j Res(f, z_j) Ind_\gamma(z_j).$$

Il teorema dei residui è utilizzato nel calcolo di particolari integrali di funzioni di variabile reale; questa applicazione non sarà trattata in questa sezione perché va oltre gli obiettivi della stessa.
Per chi fosse interessato suggeriamo ([27], §16; [16], §10).

Teorema
Se z_0 è un polo di ordine m per f ($m \geq 1$ e intero), allora

$$Res(f, z_0) = \frac{1}{(m-1)!} \lim_{z \to z_0} \frac{d^{m-1}}{dz^{m-1}} (f(z)(z - z_0)^{m-1}).$$

Questa formula è molto interessante se z_0 è un polo semplice in quanto si riduce a

$$Res(f, z_0) = \lim_{z \to z_0} \big(f(z)(z - z_0)\big).$$

4. GRAFICI DI FUNZIONI

In questa breve sezione, tratteremo del modo di rappresentare una funzione di variabile complessa mediante dei grafici. Nella sezione di richiami di analisi complessa si è visto che molti risultati sulle funzioni di variabile complessa derivano da analoghe proprietà per quelle di due variabili reali.
Noteremo che varranno analoghe similitudini anche per quanto riguarda i modi più comuni di rappresentarle mediante dei grafici.
Inizieremo, dunque, con una breve panoramica sui grafici di funzioni di due variabili reali per poi passare a quelle di variabile complessa. Saranno considerate come assodate conoscenze relative ai grafici di funzioni di una variabile reale, cioè gli usuali studi di funzione con cui si ha a che fare fin dalle scuole secondarie.
Per quanto riguarda l'analisi complessa, il discorso è piuttosto complicato e si ricorre ad espedienti non proprio ortodossi che ci consentono di avere una visione solamente parziale del comportamento di una funzione. Nelle funzioni di una e due variabili reali dal grafico si possono trarre informazioni visuali circa la continuità, i massimi e i minimi mentre per quelle di una variabile complessa usualmente si rappresenta il modulo dei valori dell'immagine per cui le informazioni che si traggono visivamente non riguardano la funzione in esame ma il suo modulo.

4.1 FUNZIONI DI DUE VARIABILI REALI

4.1.1 Grafici tridimensionali

Una funzione di due variabili reali $f(x, y)$ è un'applicazione che ad un elemento di $\mathbb{R}^2$ associa un numero reale secondo la legge espressa dalla funzione stessa.
La questione si complica, rispetto alle usuali funzioni di una variabile, proprio per via della doppia dipendenza di f da x e da y. In linea di massima basta prendere uno *schema* nel quale, associando due assi cartesiani alle altrettante variabili indipendenti, nel terzo asse avremo il valore di f in analogia a quanto accade per funzioni di una variabile.
Tuttavia, seguendo questa linea di pensiero, il grafico di una funzione di due variabili reali sarà tridimensionale. In esso sull'asse x rappresenteremo i valori delle x, sull'asse y quelli delle y mentre lungo l'asse z verranno individuati i valori di $z = f(x, y)$. (Figura 4.1).

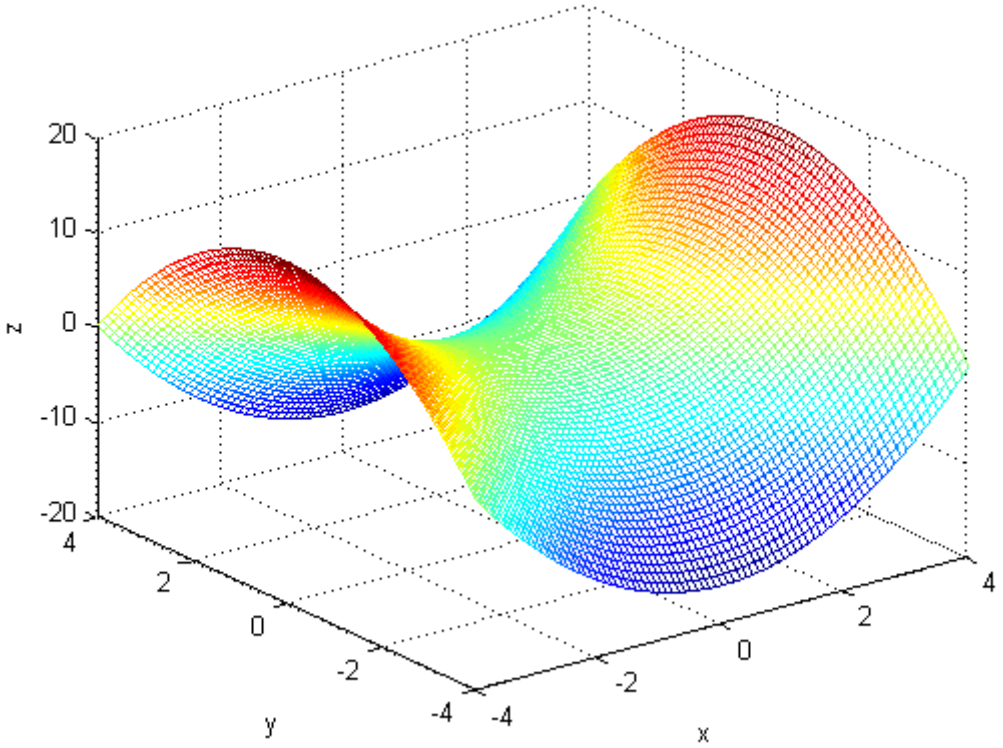

Figura 4.1. Grafico della funzione $f(x, y) = x^2 - y^2$ (matlab).

La Figura 4.1 mostra il grafico tridimensionale di $f(x, y) = x^2 - y^2$.
L'idea di base è sempre la stessa; si rappresenta la variazione di una variabile dipendente (z), in base ai valori delle variabili indipendenti (x e y) mediante la legge espressa dalla funzione. Il grafico è tridimensionale proprio perché due assi occorrono per la rappresentazione dei valori delle x e y mentre nel terzo sono mostrati i valori assunti da $z = f(x, y)$.
La funzione dell'esempio è già stata vista nella sezione di richiami di Analisi Matematica II: nell'origine ha un punto di sella, cioè un punto critico che non è né un massimo né un minimo. Si può vedere anche dal grafico la motivazione della nomenclatura "punto di sella".
Per quanto riguarda le funzioni in due variabili, il grafico tridimensionale è di facile interpretazione anche se difficile da costruire in senso pratico. L'aiuto di un computer, tramite appositi programmi, si rivela fondamentale poiché riesce a fornire in breve termine un'immagine molto precisa con la quale (in genere) si può anche interagire cambiando punto di vista o analizzando zone di interesse nel dettaglio.
Vediamo un esempio di una funzione più complessa della precedente: $f(x, y) = \cos(x^2 + y^2)$.

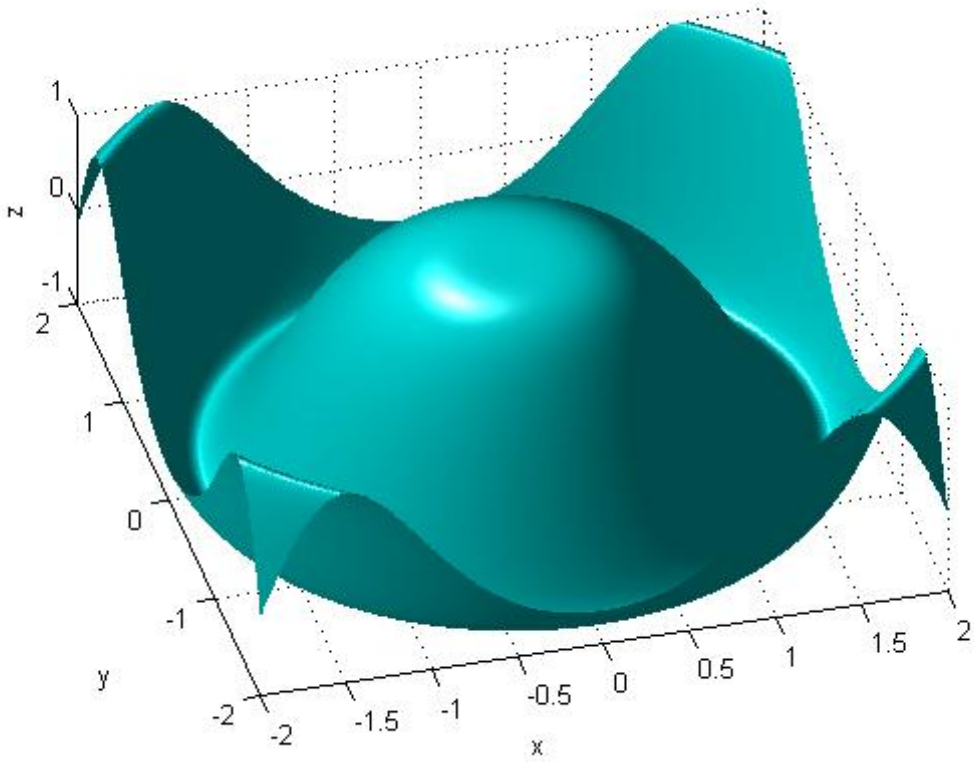

Figura 4.2. Grafico della funzione $f(x, y) = \cos(x^2 + y^2)$ (matlab).

La Figura 4.2 mostra il grafico in tre dimensioni della funzione appena citata. La differenza con la Figura 4.1 è solamente stilistica, nella prima si è scelto di rappresentare la funzione tramite una griglia color arcobaleno mentre la seconda è una superficie a tinta unita.
Entrambe le immagini sono state realizzate con l'ausilio del Matlab, un programma molto utilizzato in ambito matematico e ingegneristico poiché dispone di molti mezzi per lo studio avanzato di funzioni e modelli.
Tralasciando l'aspetto artistico, il grafico tridimensionale di una funzione a due variabili è il metodo più semplice e intuitivo per rappresentarla sebbene si riveli efficace esclusivamente con l'ausilio di un calcolatore ed appositi programmi. Il computer, infatti, presenta tre vantaggi fondamentali:

- *precisione*, il grafico è tracciato con un grado di (im)precisione prefissato dall'utente;
- *manipolazione* (ci sono programmi, ad esempio, che consentono di cambiare punto di vista nel grafico o analizzare zone specifiche senza dover ricominciare tutto daccapo);
- *tempistica*, tracciare un grafico con un programma per computer consente di avere un risultato in tempi rapidi dipendenti, generalmente, dalle caratteristiche della macchina.

Tuttavia, anche per il computer ci sono dei limiti dovuti, principalmente, alla precisione con cui si sceglie di costruire il grafico. Nell'esempio del Matlab, il grafico viene tracciato sulla base di una serie di punti disposti su una griglia scelti in base alla precisione voluta. Essi sono uniti per costruire l'immagine finale: per aumentare la precisione bisogna far crescere il numero dei punti e, dunque, dei calcoli con conseguente allungamento dei tempi di elaborazione.

4.1.2 Grafici bidimensionali

Oltre al *semplice* grafico tridimensionale, ci sono altri metodi altrettanto efficaci per rappresentare le funzioni in due variabili che si servono di tecniche grafiche di utilizzo comune soprattutto nella cartografia. Essi sono principalmente di due tipi:

- grafico a *variazione di colori*
- grafico a *curve di livello*.

A primo impatto potrebbero sembrare concetti strani e, forse, anche esotici tuttavia sono grafici comunemente utilizzati in cartografia (Figura 4.3).

Figura 4.3. Frammento dell'Italia Fisica (Atlante DeAgostini [6]*).*

La Figura 4.3 è presa dall'atlante DeAgostini ([6]). Essa è una rappresentazione bidimensionale di una realtà – nel nostro caso il centro Italia – tridimensionale. L'artificio utilizzato è quello di servirsi di varie tonalità di colore per dare l'illusione del tridimensionale oltre che per fornire informazioni riguardanti l'altitudine del territorio, o del mare, in un dato punto. Per esempio, grazie a questo artificio sappiamo che l'area all'interno del cerchio rosso nella Figura 4.3 ha un'altitudine compresa tra i 500 e i 1000 metri s.l.m. (vedere legenda a lato dell'immagine stessa).

Per le funzioni a due variabili, il discorso è analogo. Si traccia un grafico bidimensionale in cui i colori variano a seconda del valore assunto dalla funzione stessa al variare di x e y. Ovviamente tale grafico sarà corredato da un'apposita legenda che indica a quale valore corrisponde un colore specifico (Figura 4.4).

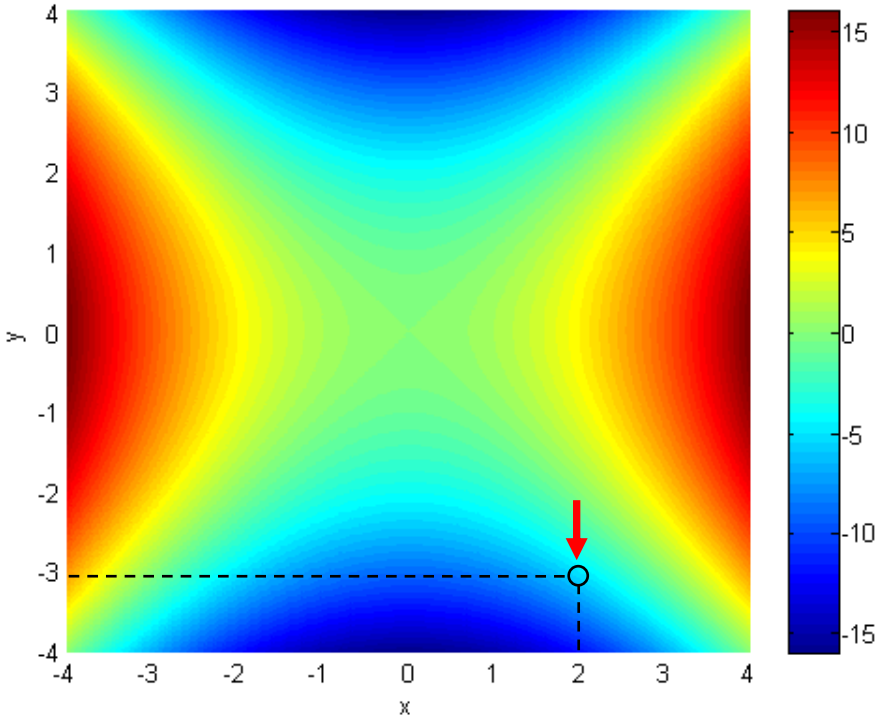

Figura 4.4. Grafico bidimensionale a colori della funzione $f(x,y) = x^2 - y^2$ (matlab).

Nell'immagine di Figura 4.4 si è voluta scegliere come funzione da rappresentare mediante il grafico bidimensionale $f(x,y) = x^2 - y^2$, la stessa della Figura 4.1. Questa scelta non è casuale poiché, in questo modo, si può fare un raffronto diretto con il grafico tridimensionale. Analogamente alla carta geografica, la scala di colori ci dice, ad esempio che il punto $(2,-3)$ – indicato dalla freccia rossa nel disegno – ha un'immagine che risulta valere, approssimativamente, -5.

Poi, ovviamente, $f(2,-3) = 2^2 - (-3)^2 = -5$ è la conferma di quanto visto dal grafico.

Da notare che anche in questa rappresentazione, sebbene non perfettamente visibile, si può notare che l'origine è un punto particolare poiché, a partire da essa, la funzione cresce ad est/ovest mentre decresce a nord/sud.

Un altro tipo di grafici è ottenuto servendosi delle curve di livello. In altre parole il grafico è un insieme di curve (distinte!) – dette *curve di livello* – nelle quali la funzione che ci interessa assume un determinato valore. Per gli altri punti che non giacciono su una di queste curve si deduce che la funzione assume un valore compreso tra le due curve che delimitano la zona nella quale si trova il punto che ci interessa. Un grafico simile si incontra comunemente in meteorologia, nelle carte che mostrano le varie aree di pressione (Figura 4.5).

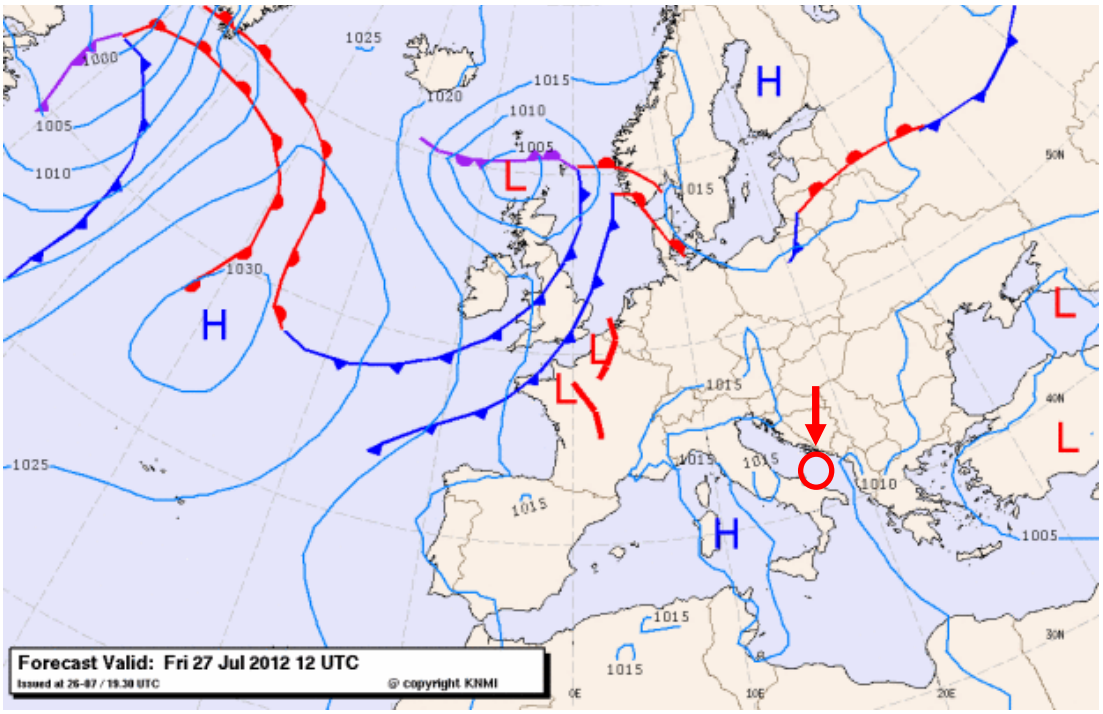

Figura 4.5. Andamento della pressione atmosferica in Europa (27-07-2012).

Nella Figura 4.5, presa dal sito www.ilmeteo.it ([32]), si può notare l'andamento della pressione atmosferica alle 12 del giorno 27-07-2012. Nella mappa sono rappresentate le curve di livello della pressione, cioè l'insieme dei punti nei quali la pressione atmosferica assume il valore costante indicato nell'etichetta (in millibar). Nella zona indicata dal cerchio, ad esempio, deduciamo che la pressione è compresa tra i 1010 e i 1015 mbar trovandosi nello spazio tra le curve di livello dei 1010 e 1015 millibar.
Per le funzioni a due variabili si può fare un discorso analogo (Figura 4.6).

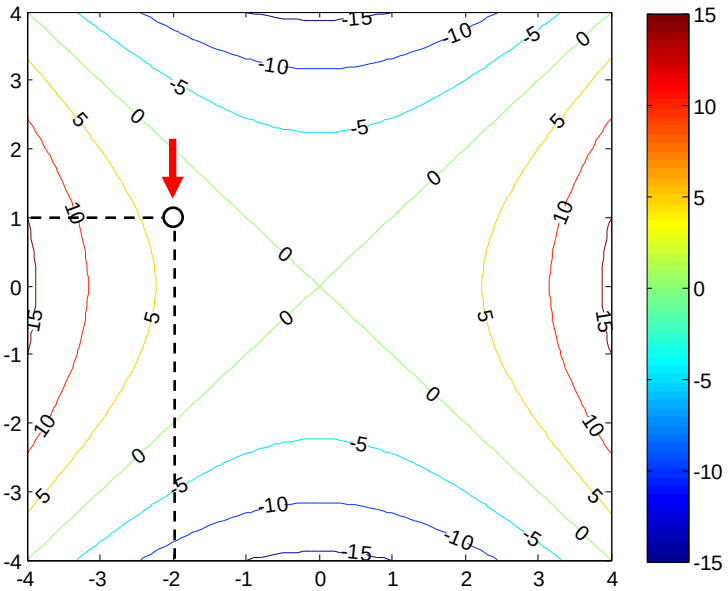

Figura 4.6. Grafico a curve di livello della funzione $f(x, y) = x^2 - y^2$ (matlab).

La Figura 4.6 è un esempio di grafico bidimensionale con curve di livello per una funzione di due variabili reali. I grafici a curve di livello sono un'esemplificazione del caso precedente.

Le curve di livello rappresentano il luogo dei punti in cui la funzione a due variabili assume lo stesso valore fissato; nel grafico esse hanno un'etichetta che rappresenta proprio il valore assunto da $f(x,y)$ nei loro punti. In termini matematici una curva di livello è

$$\gamma = \{(x,y) \in \mathbb{R}^2: \quad f(x,y) = c\}, \qquad c \in \mathbb{R}.$$

In realtà, così come per i punti, anche le curve di livello sono infinite, tuttavia se ne rappresentano solo alcune ritenute significative ai fini del grafico. I punti sulle curve di livello sono tali che la funzione assume il valore indicato dalla curva stessa mentre gli altri sono in una posizione interna tra due diverse curve e il valore assunto è, dunque, intermedio. Nell'esempio di Figura 4.6, nel punto $(-2,1)$ – indicato dalla freccia – intuiamo che la funzione assume un valore compreso tra lo 0 della curva a sinistra e il 5 della curva a destra trovandosi all'interno della regione delimitata dalle due.
Anche qui si è volutamente scelto di rappresentare il grafico della funzione $f(x,y) = x^2 - y^2$ in modo da poter raffrontare il risultato ottenuto con quello di Figure 4.4 e Figura 4.1.

4.2 GRAFICO DI UNA FUNZIONE DI VARIABILE COMPLESSA

4.2.1 Introduzione

Ci sono molti modi di rappresentare una funzione di variabile complessa mediante dei grafici, tuttavia in questa sezione ne analizzeremo solo alcuni, i più significativi per gli scopi della tesi.
I metodi proposti si rifanno a quelli per rappresentare funzioni di due variabili reali dividendo z nell'usuale scomposizione $z = x + iy$ con $x = Re(z)$ e $y = Im(z)$. Questa interpretazione è molto intuitiva, però si presta a grandi interrogativi che non riescono a trovare una risposta esauriente come nel caso di funzioni di due variabili reali.
Essi sono legati alla seguente affermazione.
"Una funzione di variabile complessa è un'applicazione che associa ad un valore $\underline{z \in \mathbb{C}}$ un corrispettivo $w = f(z)$ univocamente determinato mediante la legge espressa dalla funzione stessa e anch'esso <u>complesso</u>."
Il problema principale è proprio quello sottolineato nella definizione – un po' semplicistica ma efficace – di una funzione di variabile complessa

$$f: A \subseteq \mathbb{C} \to \mathbb{C}, \qquad z \in \mathbb{C} \longmapsto f(z) \in \mathbb{C}.$$

Nella usuale visualizzazione otteniamo $z = x + iy$ e $f(z) = u + iv$ per opportuni valori $u, v \in \mathbb{R}$. Il problema è quindi quello di rappresentare $f(z)$ al variare di z poiché per il grafico completo occorrerebbero 4 assi cartesiani: 2 per l'input e 2 per l'output.
Tuttavia, senza addentrarsi in particolari questioni di relatività generale o di geometria differenziale, lo spazio con cui siamo abituati ad interfacciarci è tridimensionale e permette, al massimo, l'utilizzo di 3 assi coordinati per rappresentare grafici nello spazio.

Si cerca, quindi, di ovviare il problema con una conseguente perdita dell'informazione.

4.2.2 Tridimensionale (modulo)

Il primo passo è il seguente.
Dopo aver compreso che il problema sta nella rappresentazione completa di una funzione di una variabile complessa, ci si chiede se questo può essere aggirato mediante qualche espediente. La risposta è negativa, ma si arriva all'idea di rappresentare, non $f(z)$ ma $|f(z)|$, il suo modulo.
Sappiamo che $\forall z \in \mathbb{C}, |z| \in \mathbb{R}$ e quindi ci ritroviamo ad avere il grafico di una funzione che, associato un valore in input $z \in \mathbb{C}$ restituisce $|f(z)| \in \mathbb{R}$ ed è quindi perfettamente equiparabile ad una funzione di due variabili reali rappresentabile con un grafico tridimensionale (Figura 4.7).

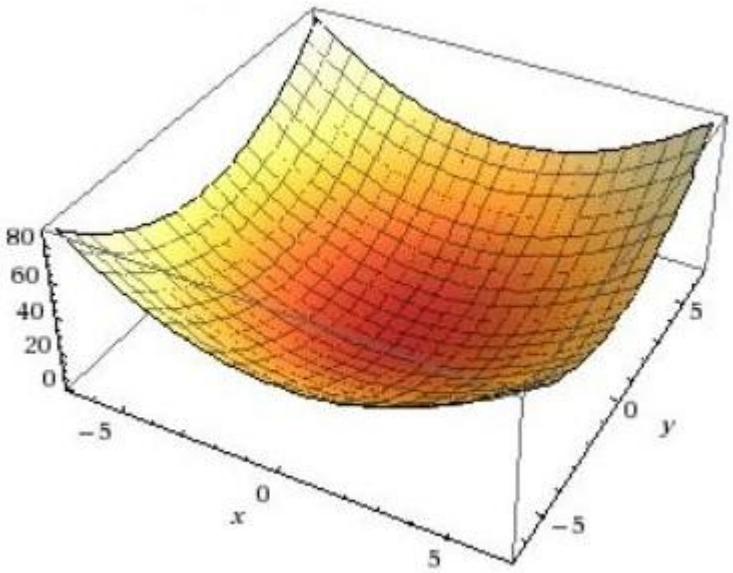

Figura 4.7. Grafico tridimensionale della funzione $f(z) = z^2$.

La Figura 4.7, realizzata tramite il sito di wolframalpha ([33]), è relativo alla funzione $f(z) = z^2$. Esso, in realtà, non rappresenta la funzione nella sua interezza, ma il suo modulo. Quindi eventuali massimi e/o minimi che si colgono in esso non sono relativi alla funzione z^2 ma al suo modulo.
E' evidente che la perdita di informazione che si ottiene restringendoci al modulo di una funzione di variabile complessa riguarda la visualizzazione dei punti critici della stessa. Tuttavia questi grafici sono molto utilizzati poiché non si perdono, invece, informazioni riguardanti gli zeri della funzione.
A tale proposito ricordiamo la seguente proprietà del modulo

$$|z| = 0 \iff z = 0.$$

4.2.3 Grafico tridimensionale ($Re(z)$ o $Im(z)$)

Accanto al grafico precedente, ce ne sono altri due degni di nota per quanto riguarda una funzione di variabile complessa: il primo consiste nel rappresentare $Re\big(f(z)\big)$ mentre il secondo $Im\big(f(z)\big)$, al variare di $z \in \mathbb{C}$.
Essi sono due grafici indipendenti che indicano, semplicemente, come variano la parte reale e immaginaria della funzione in esame al variare del valore complesso in input (Figure 4.8a e 4.8b).

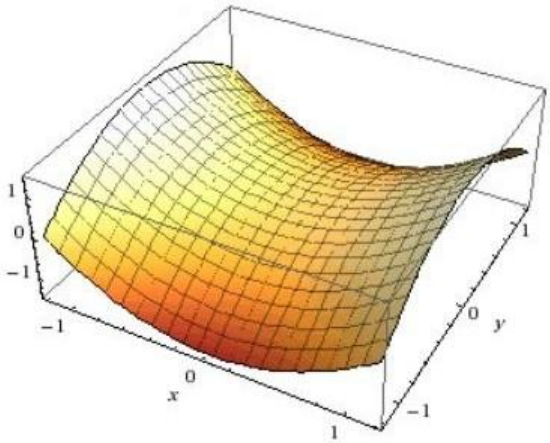

Figura 4.8a. Grafico di $Re\big(f(z)\big)$ per $f(z) = z^2$.

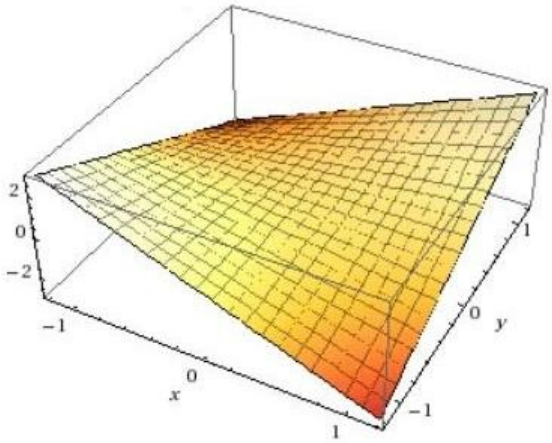

Figura 4.8a. $Im\big(f(z)\big)$ per $f(z) = z^2$.

Le Figure 4.8a e 4.8b sono state entrambe realizzate tramite il sito wolframalpha ([33]): esse si riferiscono alla rappresentazione tridimensionale, rispettivamente, di $Re(z^2)$ e $Im(z^2)$.
L'utilità di questi grafici è quella di individuare andamenti ed eventuali punti critici non della funzione in esame ma della sua parte reale e immaginaria. Per avere un quadro più completo basta individuare uno stesso punto su entrambi i grafici e vedere come si comportano singolarmente la parte reale e immaginaria della funzione per poi ottenere $f(z)$ come $f(z) = u + iv$.

4.2.4 Altri tipi di grafici (bidimensionali)

Le similitudini tra una funzione di variabile complessa e una di due variabili reali che si ottengono nel considerare $f(z) = f(x, y) = u(x, y) + iv(x, y)$ – già esaminate nella sezione di Richiami di Analisi Complessa – ci hanno consentito di creare i grafici tridimensionali visti nei paragrafi precedenti. Tuttavia, così come per le funzioni in due variabili, anche per quelle in una variabile complessa si possono considerare grafici bidimensionali tramite curve di livello o scale di colori.
La differenza sta all'origine: le curve di livello o le variazioni di colori non rappresenteranno $f(z)$ ma, a seconda dei casi, $|f(z)|, Re\big(f(z)\big)$ o $Im\big(f(z)\big)$.

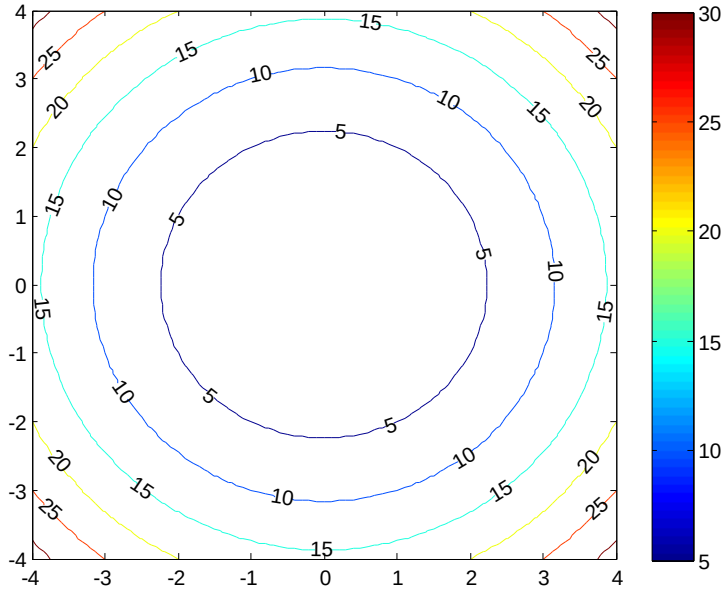

Figura 4.9. Grafico a linee di contorno per la funzione $f(z) = z^2$ (matlab).

In linea di principio, la didascalia della Figura 4.9 è sbagliata. Le linee di contorno, anche se riferite alla funzione $f(z) = z^2$ sono, in realtà, del modulo della stessa e, cioè, $|f(z)| = |z^2|$. Si può trovare un riscontro tra le curve di livello della Figura 4.9 e il corrispettivo grafico tridimensionale di Figura 4.7 per la funzione $f(z) = z^2$.

Un discorso analogo vale anche per i grafici a curve di livello di $Re\big(f(z)\big), Im\big(f(z)\big)$ e, in generale, si può estendere a quelli a variazione di colore come trattato nei paragrafi precedenti nel caso di funzioni di due variabili reali.

5. TEORIA DEI NUMERI – DIVISIBILITA', NUMERI PRIMI E CONGRUENZE

In questa sezione analizzeremo argomenti basilari di quella grande branca della Matematica rappresentata dalla Teoria dei Numeri. Inizieremo definendo la divisibilità tra interi per poi passare ai numeri primi ed infine accennare alle congruenze.

5.1 DIVISIBILITA' E NUMERI PRIMI

5.1.1 Introduzione

> <<La Matematica è la regina delle scienze e la Teoria dei Numeri è la regina della Matematica.>>
>
> C. F. Gauss (1777-1855)

Molti testi di Teoria dei numeri iniziano citando questa famosa affermazione del matematico Gauss. Essa non è la solita sentenza di un amante della materia ma sintetizza un pensiero comune tra i matematici: la Teoria dei Numeri, nella sua apparente semplicità, racchiude alcuni dei misteri più difficili e appassionanti dell'intera Matematica.
La Teoria dei Numeri – abbreviato in TDN – si interessa principalmente delle proprietà degli interi (insieme $\mathbb{Z}$) a proposito di divisibilità, primalità ed altre nozioni collegate che ci capita di incontrare fin dalle scuole elementari. Tuttavia a partire da premesse così innocenti si finiscono per raggiungere orizzonti impensabili, che coinvolgono l'Analisi e l'Algebra e vanno a costituire una branca della Matematica molto più complessa e affascinante di quello che potrebbe far venire in mente la *descrizione ufficiale*.
In quest'ottica possiamo affermare che la TDN è senz'altro la più antica ma anche una tra le più moderne e attuali tra le Matematiche.

- E' la più antica perché, fondamentalmente, nasce con la necessità dell'uomo di contare e di operare le operazioni elementari con i numeri.
- E' una delle più moderne poiché molti problemi matematici ancora irrisolti – tra i quali alcuni famosi problemi del millennio ([22]) – traggono origine dai suoi misteriosi risvolti.
- E' senz'altro attuale e, con la parola *attuale*, non intendiamo solamente la vita di tutti i giorni che ci mette faccia a faccia con le operazioni elementari e le congruenze (vedremo che frasi come "le tre del pomeriggio" sono, in realtà, delle congruenze). I

misteri che circondano i numeri primi, ad esempio, sono alla base della moderna sicurezza informatica (internet, codici bancomat,…).

L'oggetto di questa tesi – l'Ipotesi di Riemann – nasce e si sviluppa proprio con la TDN, sebbene si snodi nei meandri dell'Analisi Matematica e, in particolar modo, dell'Analisi Complessa. Inoltre, un'eventuale conferma o smentita di questa che per ora è proprio un'ipotesi porterebbe con sé risultati importanti nella Teoria dei Numeri.

In questa tesi, cercheremo di trattare in maniera semplice ed efficace solo alcuni concetti della TDN che serviranno da background per le sezioni future. Tuttavia, per chi può vantare conoscenze matematiche al livello di Analisi Matematica I o di Liceo scientifico, varrebbe la pena approfondire questa affascinante materia che invece è spesso trascurata dai corsi specifici di molte università.

A chi volesse intraprendere un viaggio nella Teoria dei Numeri, tra gli innumerevoli testi, consigliamo i seguenti, utilizzati anche qui come riferimenti.

- *An Introduction to the Theory of Numbers*, di Hardy e Wright ([10]). E' il "classico dei classici" e approfondisce ampiamente tutte le problematiche della TDN. Tuttavia, per alcune di queste, sono indispensabili conoscenze più avanzate di vari ambiti della Matematica come Analisi e Algebra.
- *Numeri e Crittografia*, di Leonesi e Toffalori ([11]). La sezione dedicata alla TDN è molto semplice e ampiamente spiegata in tutte le sue sfaccettature; inoltre il libro è dedicato alle applicazioni della Teoria dei Numeri alla Crittografia (come suggerisce il titolo).
- *The New Book of Prime Number Records*, di Ribenboim ([17]). Un testo semplice, immediato e completo che si focalizza soprattutto nelle questioni inerenti i numeri primi.

Per chi, invece, può vantare conoscenze più avanzate di Matematica, si consiglia la lettura del *Introduction to Analytic Number Theory* di Apostol ([3]) che mostrerà come, in molti ambiti della TDN, confluiscano diverse discipline apparentemente lontane come l'Analisi.

5.1.2 Divisibilità e divisione tra interi

Divisibilità e divisione sono due concetti ben diversi, seppur strettamente legati tra loro.; il nostro punto di partenza sarà il seguente risultato ([11], §2.1).

Teorema

Siano $a, b \in \mathbb{Z}$ con $b \neq 0$. Allora esistono $q \in \mathbb{Z}$ e $r \in \mathbb{N}$ unici tali che

$$a = b \cdot q + r,$$

con $0 \leq r < |b|$.

In questo caso q si dice quoziente mentre r è il resto della divisione di a per b.

Siano, ora, $a, b \in \mathbb{Z}$ con $b \neq 0$. Diremo che b divide a o che a è divisibile per b – e scriveremo $b|a$ – se esiste un intero q tale che $a = b \cdot q$. In questo caso si dice anche che a è un multiplo di b o anche che b è un divisore di a ([23], §1.2; [10], §1.1).

La definizione appena data introduce la relazione di divisibilità. Riferendoci al teorema precedente, possiamo anche affermare che $b|a$ quando il resto r è nullo. Se, invece, b non divide a, scriveremo $b \nmid a$.
Nel caso in cui $b|a$, la quantità q può essere trovata operando la classica divisione di a per b. La divisibilità, dunque, è una relazione binaria tra interi mentre la divisione è un'operazione binaria che accompagna tale relazione, consentendoci di verificarla o smentirla.
Se abbiamo ad esempio 25 e 50, fare $50:25$ ci darà come risultato 2, senza resto. Questo vuol dire che $50 = 2 \cdot 25$ e, dunque, $25|50$. Tuttavia se avessimo 25 e 51, $51 = 25 \cdot 2 + 1$ e quindi $25 \nmid 51$.
Siano, dunque, $a, b \in \mathbb{Z}$ con $b \neq 0$. La divisibilità, intesa come relazione tra interi, gode delle seguenti proprietà ([3], §1.2; [23], §1.2), valide per ogni scelta di a, a', b, c interi.

(i) $1|a$.
(ii) Proprietà riflessiva: $a|a$.
(iii) Proprietà transitiva: $c|b, b|a$ implica $c|a$.
(iv) Linearità: $b|a, b|a'$ implica $b|(ma \pm na')$ con $a', m, n \in \mathbb{Z}$.
(v) Moltiplicazione: $b|a$ equivale a $nb|na, n \in \mathbb{Z}$.
(vi) Legge di cancellazione: $nb|na$ con $n \in \mathbb{Z}$ e $n \neq 0$ implica $b|a$.
(vii) $a|0$.
(viii) $0|a$ implica $a = 0$.
(ix) $b|a$ e $a|b$ implica $|a| = |b|$.
(x) $b|a$ e $a \neq 0$ implica $|b| \leq |a|$.
(xi) $b|a$ implica $(a/b)|a$, nel quale a/b denota il quoziente esatto della divisione tra a e b.

Tra tutte queste proprietà, alcune immediate (come la (i) ad esempio), altre un po' meno (la (x) ad esempio), degna di nota è la (iv). Molti risultati sulla divisibilità e sui numeri primi fanno leva proprio su di essa.
Con un occhio di riguardo alla (ix) e (x), diremo che b è un divisore "proprio" di a se $|b| < |a|$, cioè $a = b \cdot c$ con $c \in \mathbb{Z}$ e $|c| \neq 1$. In caso contrario, cioè se $|b| = |a|$, b è un divisore "improprio" di a; in particolare a è un divisore improprio di sé stesso. Per i naturali, la questione si riduce a dire che b è un divisore improprio di a se $b = a$ mentre non lo è se $b < a$.
Dalle (i) e (ii), possiamo osservare che un qualsiasi naturale $a \neq 0$ ha come divisori 1 e a: essi sono detti divisori banali di a ([23], §1.2). Gli eventuali altri divisori vengono chiamati divisori non banali (di a).
Diremo allora che un numero naturale n è primo se $n > 1$ e n possiede solamente divisori banali, cioè se è divisibile solo per 1 e n. In caso contrario diremo che n è composto cioè se possiede anche divisori non banali quindi $n = d \cdot d'$ con $d, d' \in \mathbb{N}$ e $1 < d, d' < n$. Diremo poi che un intero è primo o composto se tale è il suo valore assoluto.
Le definizioni e i risultati che seguono sono enunciati per numeri naturali ma si estendono in modo ovvio agli interi, con le opportune modifiche, sulla base della precedente definizione.

Teorema (fondamentale dell'aritmetica) ([10], §1.3; [11], §2.3; [23], §1.2)
Qualunque naturale $a \neq 0,1$ si decompone in uno e un solo modo – a meno dell'ordine dei fattori – come prodotto di numeri primi.

Grazie al teorema fondamentale dell'aritmetica, possiamo esprimere ogni intero $n > 1$ nella forma

$$n = \prod_{i=1}^{r} p_i^{a_i},$$

dove i vari p_i sono i fattori primi distinti che dividono n e per ogni i, $a_i \geq 1$ è l'esponente massimo con cui p_i divide n. Per esempio 12 si decompone come $2^2 \cdot 3$ $(3 = 3^1)$.
In generale (vedi [5], pag.18) si può usare la seguente notazione per tutti i naturali $n \geq 1$:

$$n = \prod_{i=1}^{\infty} p_i^{a_i}, \qquad a_i \geq 0, \qquad p_i \text{ primo}.$$

Stavolta i p_1 costituiscono la collezione (che tra poco vedremo essere infinita) di tutti i primi. La rappresentazione può sembrare ridondante e fuorviante, tuttavia, in essa, basta porre $a_i = 0$ per tutti quei fattori primi che non compaiono esplicitamente nella decomposizione di n. Per fare un esempio, prendiamo $n = 175$:

$$175 = 2^0 \cdot 3^0 \cdot 5^2 \cdot 7^0 \cdot 11^1 \cdot 13^0 \cdot 17^0 \cdot \ldots = 5^2 \cdot 11.$$

L'ultima rappresentazione non è quindi dissimile da quella che la precede: sono due modi equivalenti di vedere la decomposizione di un naturale in fattori primi.

Teorema ([3], §1.5)
Dato $n = \prod_{i=1}^{r} p_i^{a_i}$, l'insieme dei divisori positivi di n è l'insieme dei numeri della forma $\prod_{i=1}^{r} p_i^{c_i}$, dove $0 \leq c_i \leq a_i$ per $i = 1, \ldots, n$.

Corollario ([21], §2.1)
Il minimo divisore positivo $d > 1$ di un intero $n > 1$ (eventualmente coincidente con n) è primo.

Corollario ([23], §1.2)
Se $n > 2$ è un numero composto, allora n ha un divisore p primo tale che $p \leq \sqrt{n}$.

Questo risultato è molto più importante di quello che sembra e lo richiameremo spesso nella sezione dedicata al riconoscimento dei numeri primi. Inoltre poiché

$$\max\{p : p|n, p \text{ primo}\} \leq n$$

possiamo confermare che il precedente prodotto è finito:

$$n = \prod_{i=1}^{\infty} p_i^{a_i} = \prod_{i=1}^{s} p_i^{a_i}.$$

Infatti esiste $j \in \mathbb{N}$ tale che $a_k = 0$, per ogni $k \geq j$ proprio perché n non può avere un divisore primo maggiore di n stesso.

Teorema ([21], §2.1)
Se $p | \prod_{i=1}^{n} p_i$, con $p, p_1, \ldots, p_n$ primi, allora $p = p_k$ per un certo indice $k \in \{1, \ldots, n\}$.

5.1.3 La successione dei numeri primi

Si è detto che un numero > 1 è primo quando possiede solo divisori banali. Fin dall'antichità sono stati sviluppati metodi – più o meno efficaci – per stabilire la primalità di un numero. Molte illustri menti, inoltre, si sono sforzate di dare una risposta al seguente quesito:

"si può dare un ordine logico alla successione dei numeri primi?"

La successione dei numeri primi, ovvero $(p_n)_{n\in\mathbb{N}}$ con p_n primo $\forall n$, sembra illogica e priva di un ordine matematico di un qualsiasi tipo. Consideriamo comunque la sequenza

$$2,3,5,7,11,13,17,19,23,29,31,37,41,43,47$$

rappresenta i numeri primi compresi tra 1 e 50. A partire da essa si possono fare le seguenti osservazioni piuttosto elementari.

- 2 è l'unico primo pari. Tutti gli altri numeri pari, infatti, sono divisibili per 2.
- Le coppie di primi della forma $(p, p+2)$, cioè $(3,5)$ così come $(5,7)$ fino a $(41,43)$ per quanto riguarda quelli fino a 50, sono dette "primi gemelli". La nomenclatura "gemelli" sta proprio nel fatto che 2 è la distanza minima che può esserci tra due numeri primi (a parte la coppia $(2,3)$ che è l'unica eccezione). Queste coppie sono sempre state oggetto di curiosità fin dall'antichità e domande come "quanti ce ne sono" o "come stabilire se, dato p, $p+2$ sia primo o meno" sono entrate da molti secoli nei problemi (irrisolti) della TDN.

Riguardo ai primi, ci sono molte sequenze particolari degne di nota che verranno analizzate più nel dettaglio nella sezione dedicata proprio al problema di generare numeri primi. Tutte queste sequenze particolari sono il frutto della curiosità umana nella comprensione della logica – ammesso che ci sia – che si cela dietro quella successione apparentemente scoordinata che è quella dei primi.
Tuttavia, sebbene non si sappia se si celi un ordine dietro tale sequenza, ci sono altrettanti risultati sui primi che sono noti da secoli. In questo paragrafo ne vedremo due, mentre degli altri si parlerà nella prossima sezione.

Teorema (Euclide)
I numeri primi sono infiniti.

Questo teorema è noto fin dall'antichità, così come la sua dimostrazione più semplice che è proprio quella di Euclide. Tuttavia nel corso dei secoli molti altri matematici hanno trovato altre dimostrazioni, alcune eleganti come quella di Eulero ([11], §2.4; [17], §1.3), altre più sofisticate ([17], §1.6). Per chi è interessato si rimanda al testo di Ribenboim, nel quale si trovano 13 dimostrazioni suddivise in 8 categorie ([17], §1.1-1.7). Per quanto riguarda i numeri primi gemelli, invece, è ancora un problema aperto stabilire se ce ne siano infiniti o meno ([10], §1.4).

5.1.4 Massimo comun divisore e minimo comune multiplo

Consideriamo a, b due interi positivi. Diremo che un naturale $d > 1$ è il massimo comun divisore di a e b se:

- $d|a$ e $d|b$, cioè d divide tanto a quanto b;
- ogni altro divisore comune di a e b è $\leq d$.

Il massimo comun divisore di due interi positivi a e b è indicato, generalmente, con la notazione (a, b) anche se in qualche testo si può trovare la scrittura $MCD(a, b)$ o anche $\gcd(a, b)$ (dall'inglese "greatest common divisor").
La sua esistenza è garantita dal fatto che i divisori comuni di a e b includono 1 e sono un numero finito. La sua unicità scende banalmente dalla definizione.
Possiamo osservare che se a e b sono due numeri primi differenti, risulta $(a, b) = 1$; tuttavia $(a, b) = 1$, anche in altri casi particolari. In base a questa osservazione, diremo che due interi positivi a, b sono primi tra loro se $(a, b) = 1$.
Possiamo dare qualche proprietà.

- $(n, n + 1) = 1$, per qualsiasi n intero positivo.
  Infatti se esistesse $d > 1$ tale che $(n, n + 1) = d$, questo vorrebbe dire $d|n$ e $d|(n + 1)$ e quindi per la proprietà (iv) della divisibilità, $d\big|\big((n + 1) - n\big)$, cioè $d|1$ che è impossibile. Questo semplice ragionamento sta alla base della dimostrazione del teorema di Euclide sull'infinità dei numeri primi (vedi ([11], §2.4), per esempio).
- $(n, n + 2) = 1$, per n dispari.
  Il ragionamento è identico al precedente ricordandosi alla fine che $d|2$ è impossibile perché $2 \nmid n$ e $2 \nmid (n + 2)$ proprio perché n e $n + 2$ sono dispari.
- $(p, q) = 1$ per p, q due distinti primi.
- $(a, p) = 1$ per a intero positivo e p primo tale che $p \nmid a$.
- $(a, 1) = (1, a) = 1$ per qualsiasi intero positivo a.

Consideriamo, ora, due interi positivi n, m. Un naturale l si dice minimo comune multiplo di m, n se $l \neq 0$ e inoltre:

- l è multiplo tanto di a quanto di b (cioè $a|l$ e $b|l$);
- ogni altro multiplo comune di a e b è $\geq l$.

Il minimo comune multiplo di due interi positivi a, b lo indicheremo con $[a, b]$ anche se in alcuni testi si può trovare la scrittura $lcm(a, b)$ (inglese) o $mcm(a, b)$ (italiano).
La sua esistenza è garantita dal fatto che esiste almeno un multiplo comune non nullo tra a e b, ovvero $a \cdot b$, e di conseguenza un minimo multiplo comune. L'unicità segue, ovviamente, dalla definizione.

Teorema ([23], §1.2)
Siano dati due interi positivi

$$a = \prod_{i=1}^{r} p_i^{a_i}, \qquad a_i \geq 0, \qquad p_i \text{ primo}$$

e

$$b = \prod_{i=1}^{r} p_i^{b_i}, \qquad b_i \geq 0, \qquad p_i \text{ primo}.$$

Allora

$$(a, b) = \prod_{i=1}^{r} p_i^{\gamma_i}, \qquad \gamma_i \geq 0$$

e

$$[a, b] = \prod_{i=1}^{r} p_i^{\delta_i}, \qquad \delta_i \geq 0$$

dove $\gamma_i = \min\{a_i, b_i\}$, mentre $\delta_i = \max\{a_i, b_i\}$.

Corollario
Per a, b interi positivi, vale la seguente relazione

$$(a, b) = \frac{a \cdot b}{[a, b]}.$$

Quest'ultimo risultato ci dice, in maniera piuttosto banale, che il calcolo del MCD è semplice se si conosce il mcm e viceversa. Da questa formula segue che $[a, b] = a \cdot b$ se e solo se a e b sono primi tra loro. In particolare $[p, q] = p \cdot q$, per p, q numeri primi distinti.
Inoltre, $[a, 1] = [1, a] = a$ per qualsiasi a intero positivo.

5.1.5 Calcolo del MCD e del mcm

Dopo aver introdotto le nozioni e le proprietà del massimo comun divisore e del minimo comune multiplo tra una coppia di interi positivi, è lecito chiedersi – in senso pratico – se esiste e, in caso affermativo, come implementare un algoritmo che ci consenta il calcolo di queste due quantità.
Un grande aiuto ci giunge dal corollario appena visto, infatti la relazione

$$(a, b) = \frac{a \cdot b}{[a, b]}$$

ci consente di calcolare solamente una delle due quantità per poi trovare facilmente l'altra. Quindi la questione si riduce al calcolo di *uno solo* tra MCD e mcm.
In realtà esistono due algoritmi molto semplici per il calcolo sia del massimo comun divisore sia del minimo comune multiplo: a tal proposito consideriamo $a, b > 1$ e interi.

- Calcoliamo (a, b).
  Scomponiamo a e b nel prodotto dei fattori primi e, servendoci del teorema precedente otteniamo facilmente il MCD.
  Vediamo di fare un esempio, siano $a = 35$ e $b = 56$. Otteniamo
  $$35 = 5 \cdot 7, \qquad 56 = 2^3 \cdot 7.$$
  Per avere $(35,56)$, dal teorema precedente otteniamo
  $$(35,56) = 7,$$

proprio perché occorre prendere i fattori comuni con l'esponente più basso.

- Calcoliamo $[a, b]$.

  Possiamo procedere nel modo precedente applicando l'altra implicazione del teorema o, in alternativa, rispolverare un simpatico metodo appreso alle scuole medie che consiste nell'elencare i multipli di entrambi i numeri fino ad ottenere il più piccolo in comune.

  Prendiamo nuovamente $a = 35$, $b = 56$.

  $35, \quad 70, \quad 105, \quad 140, \quad 175, \quad 210, \quad 245, \quad 280, \quad 315, \quad \ldots$

  $56, \quad 112, \quad 168, \quad 224, \quad 280, \quad \ldots$

  Il mcm è 280. Se, invece, volevamo applicare il metodo precedente, bastava scomporre i due numeri nel prodotto di fattori primi

  $$35 = 5 \cdot 7, \qquad 56 = 2^3 \cdot 7,$$

  prendendo tutti i fattori che compaiono (anche non comuni) con l'esponente più grande, in accordo al teorema già citato. In questo caso

  $$[35,56] = 2^3 \cdot 5 \cdot 7 = 280.$$

Prima di andare avanti sono opportune alcune osservazioni.

Questi metodi sono semplici, ma si basano sulla fattorizzazione che in realtà è un problema molto più complesso di quello che sembra e verrà trattata in maniera più approfondita nella sezione dedicata ai numeri primi. In linea di massima, però, si capisce facilmente che scomporre un numero nei suoi fattori primi è un'operazione che cresce di difficoltà al crescere del numero stesso. Scomporre 56 – con le tabelline alle elementari avremmo detto $8 \cdot 7 = 56$ – è molto più semplice che scomporre, ad esempio, il numero 8128 per il quale occorre operare più divisioni.

Possiamo, inoltre, osservare che è verificata la relazione del corollario

$$(35,56) = \frac{35 \cdot 56}{[35,56]},$$

cioè

$$7 = \frac{1960}{280}.$$

Esiste, tuttavia, un metodo molto più rapido che si svincola dalla fattorizzazione ed è l'algoritmo euclideo delle divisioni successive ([11], §2.1).

Esso si basa sulla seguente osservazione.

Siano a, b due naturali non nulli e $a \geq b$.

Operiamo la divisione tra a e b ottenendo un quoziente q e, eventualmente, un resto r.

$$a = b \cdot q + r$$

Se $r = 0$ concludiamo che a è un multiplo di b, dunque $(a, b) = a$. Consideriamo, quindi, il caso $r \neq 0$ per andare avanti con la nostra osservazione.

Il MCD tra a e b, per definizione, è un numero che divide tanto a quanto b dunque concludiamo che esso deve dividere anche r per mantenere l'uguaglianza, cioè

$$(a, b) = (b, r).$$

Questa osservazione è alla base dell'algoritmo euclideo delle divisioni successive che ora andremo a esaminare nel dettaglio.

Algoritmo euclideo

L'algoritmo euclideo – a dispetto della sua *età* – è senza dubbio il migliore per il calcolo del MCD tra due interi positivi. Inoltre esso consente di calcolare brevemente anche il mcm ricordando la relazione

$$(a,b) = \frac{a \cdot b}{[a,b]}.$$

Abbiamo, dunque, due interi positivi a, b con $a \geq b$ e $a, b \neq 0$. In maniera banale, se $a = b$, concludiamo $(a,b) = a, b$ senza nemmeno andare avanti con l'algoritmo.

Operiamo, dunque, la divisione di a per b:

$$a = b \cdot q_0 + r_0, \qquad 0 \leq r_0 < b.$$

Se $r_0 = 0$, come abbiamo già visto nell'osservazione precedente, a è un multiplo di b e l'algoritmo termina dando b come risultato. In alternativa abbiamo $(a,b) = (b, r_0)$ e iteriamo il passo appena visto dividendo b per r_0 e operando gli stessi analoghi ragionamenti.

$$b = r_0 \cdot q_1 + r_1, \qquad 0 \leq r_1 < r_0.$$

Se $r_1 = 0$, allora $(b, r_0) = r_0$ mentre in caso contrario andiamo avanti dividendo r_0 per r_1.

Dopo un numero finito di passi – che supponiamo essere $s > 0$ (s intero) – l'algoritmo avrà termine ottenendo, al passo s

$$r_{s-2} = r_{s-1} q_s, \qquad r_s = 0$$

e di conseguenza

$$(a,b) = (b, r_0) = \cdots = (r_{s-2}, r_{s-1}) = r_{s-1}.$$

Questo algoritmo è semplice e ci consente di trovare il MCD tra due interi positivi in un numero relativamente breve di passi.

Se, ad esempio, avessimo $a = 720$ e $b = 112$ e volessimo calcolare $(720,112)$:

- $720 = 112 \cdot 6 + 48$, al primo passo;
- $112 = 48 \cdot 2 + 16$, al secondo passo (si itera la procedura);
- $48 = 16 \cdot 3$, al terzo passo.

Concludiamo, dunque, che $(720,112) = 16$.

5.2 CONGRUENZE

In questa sottosezione tratteremo delle congruenze, un argomento più vicino di quanto si possa immaginare alla vita di tutti i giorni. Sarà una trattazione indolore che si focalizzerà sul concetto di relazione di congruenza e sulle proprietà di questa senza addentrarsi in sentieri impervi che conducono diritti all'algebra.

Dalle congruenze ai campi, infatti, il passo è breve (si veda, ad esempio, ([11], §3.7) e sgg.).

5.2.1 La relazione di congruenza

Introduciamo, innanzitutto, la relazione di congruenza modulo n.
Diremo che due interi a, b sono congrui modulo n – e scriveremo $a \equiv b \pmod{n}$ – se e solo se n divide $a - b$.
Il concetto in sé è molto semplice: consideriamo $a = 15, b = 3$ e $n = 12$.
Poiché $12|(15 - 3)$ allora possiamo dire che $15 \equiv 3 \pmod{12}$.
L'esempio non è preso a caso. Ripetiamo, infatti, che le congruenze sono parte integrante della vita quotidiana e di espressioni usuali come "sono le 3 del pomeriggio". E' equivalente dire "sono le 15" oppure "sono le 3 del pomeriggio" allo stesso modo di "$15 \equiv 3 \pmod{12}$". Con il termine "pomeriggio", infatti, indichiamo la parte di giornata che inizia dopo mezzogiorno – cioè le 12 – quindi l'analogia con le congruenze è immediata proprio perché dopo mezzogiorno il conteggio delle ore "può" ricominciare daccapo fino alla mezzanotte, e via dicendo.
La parte di TDN che tratta delle congruenze è detta aritmetica modulare o anche aritmetica dell'orologio in riferimento a quanto è stato appena detto.
Sebbene il quadrante dell'orologio sia l'esempio più lampante di congruenza, nella vita quotidiana di congruenze ce ne sono tante altre. Ne proponiamo alcune tra le più importanti ([10], §5.2) oltre quella riguardante le ore antimeridiane e pomeridiane.

- "Ci vediamo alle 15 di domani". Le 15 di domani, a parte il giorno, sono equivalenti alle 15 di oggi: questo è un esempio di congruenza modulo 24 che possiamo scrivere come "$15 + 24 \equiv 15 \pmod{24}$. Attenzione a non sottovalutare questa scrittura che sarà ripresa nel prossimo paragrafo.
- "Oggi è giovedì". Questo è un esempio di congruenza modulo 7: giovedì prossimo sarà giovedì proprio come oggi, ma con la differenza di una settimana. Anche "il 23 Agosto" è una congruenza, in questo caso modulo 365. Il 23 Agosto 2012 e il 23 Agosto 2013 sono lo stesso giorno a distanza di un anno, cioè 365 giorni.
- "Il dodicesimo titolare". Una frase usata non di rado nel calcio che sta ad indicare la prima riserva che scenderà nel rettangolo di gioco a partita in corso. I titolari sono undici, così dire "dodicesimo titolare" equivale a dire "la prima riserva" ($12 \equiv 1 \pmod{11}$).

Gli esempi sulle congruenze, dunque, abbondano. Abbiamo scelto quelli più immediati per sottolineare quanto sia intuitivo questo concetto.

5.2.2 Un punto di vista differente sulle congruenze

Nel paragrafo precedente abbiamo introdotto la relazione di congruenza. In generale si suppone $n \geq 2$ poiché il caso $n = 1$ ha poco significato ($1|(a - b)$, per ogni a, b).
Tuttavia, un'analisi più approfondita mostra che la relazione di congruenza $\equiv$ è anche una relazione di equivalenza. Valgono infatti le seguenti proprietà ([3], §5.1):

- proprietà riflessiva, cioè $a \equiv a \pmod{n}$, per ogni $a \in \mathbb{Z}, n \geq 2$;
- proprietà simmetrica, $a \equiv b \pmod{n}$ implica $b \equiv a \pmod{n}$, per ogni $a, b \in \mathbb{Z}$;
- proprietà transitiva, $a \equiv b \pmod{n}, b \equiv c \pmod{n}$ implicano $a \equiv c \pmod{n}$, per ogni scelta di a, b, c interi.

Come per ogni relazione di equivalenza, possiamo considerare anche per la congruenza modulo n il corrispondente insieme quoziente di $\mathbb{Z}$. Questo sarà l'insieme delle classi di resto o classi di congruenza modulo n. La classe di congruenza modulo n di un intero a si indicherà

$$a_n = \{b \in \mathbb{Z}: a \equiv b \pmod{n}\}$$

e si chiamerà anche la classe di resto di a modulo n. L'insieme quoziente delle classi di resto modulo n è indicato con $\mathbb{Z}_n$.

In realtà è facile notare che a è sempre congruente al suo resto nella divisione per n e che due interi compresi tra 0 e n (escluso) sono congruenti modulo n se e solo se sono uguali. Ne segue che gli elementi distinti di $\mathbb{Z}_n$ sono le classi a_n con $0 \leq a < n$.

Vediamo di fare un esempio e scegliamo $n = 12$. La scrittura

$$7_{12}$$

indica proprio la classe resto 7 modulo 12, ovvero

$$7_{12} = \{b \in \mathbb{Z}: b \equiv 7 \pmod{12}\} = \{\ldots, -29 - 17 - 5,7,19,31, \ldots\} = \{12 \cdot b + 7, b \in \mathbb{Z}\}.$$

5.2.3 Operazioni con le congruenze

Osserviamo che, per ogni $a, b \in \mathbb{Z}$, se $a \equiv a' \pmod{n}$ e $b \equiv b' \pmod{n}$, allora

- $a \pm b \equiv (a' \pm b') \pmod{n}$;
- $a \cdot b \equiv a'b' \pmod{n}$.

Allora possiamo definire due operazioni (di addizione e moltiplicazione) in $\mathbb{Z}_n$ per ogni scelta di interi a e b:

- $a_n + b_n = (a + b)_n$;
- $a_n \cdot b_n = (a \cdot b)_n$.

Si verifica che, rispetto a queste operazioni, $\mathbb{Z}_n$ diventa un anello commutativo unitario. Quando n è primo allora è addirittura un campo in quanto ogni elemento (escluso lo zero) ammette un inverso come vedremo nella prossima sezione parlando di equazioni congruenziali (§6.2.8). Quando n è composto, invece, esistono i così detti divisori dello zero (cioè quantità non nulle che, moltiplicate tra loro, sono congruenti a zero modulo n).

Le potenze modulo n si introducono di conseguenza, come per gli interi. Ad esempio per a intero e m intero positivo, $(a_n)^m$ è il prodotto di m fattori tutti uguali ad a_n. Per il relativo calcolo ci si può però affidare a un metodo dovuto a Legendre che consente di accelerare la procedura. Il metodo di Legendre si applica in generale, ma tra le classi di resto modulo n diventa ancora più utile, visto il contesto finito e la possibilità di ridurre ogni potenza tra 0 e n al costo di una divisione.

Si distingue se m è pari oppure no.

- Se m è pari, si ha $m = 2k$ con k naturale, e quindi per ogni intero a, vale $a^m \equiv (a^k)^2 \pmod{m}$.
- Se invece m è dispari, e dunque $m = 2k + 1$ con k naturale, e quindi per ogni intero a, vale $a^m \equiv (a^k)^2 \cdot a \pmod{m}$.

Utilizzando queste osservazioni, l'elevamento alla potenza m si riduce a una successione (non eccessivamente lunga) delle due operazioni "quadrare" e "moltiplicare per a" (modulo n, naturalmente).
Il seguente passo tratto dal libro di DuSautoy, *L'enigma dei numeri primi* ([8], §2), celebra le singolari proprietà dell'aritmetica dell'orologio, quelle che abbiamo ricordato e altre ancora.

> <<Uno dei maggiori fra i primi contributi matematici di Gauss fu l'invenzione del "calcolatore a orologio". Non si trattava di una macchina materiale, ma di un'idea che apriva la possibilità di fare aritmetica con numeri che in precedenza erano stati considerati troppo ingombranti. Il calcolatore a orologio funziona in base all'identico principio di un orologio convenzionale. Se il vostro orologio dice che sono le 9 e voi aggiungete 4 ore, la lancetta delle ore si sposterà sull'una. Allo stesso modo, il calcolatore a orologio di Gauss fornirebbe 1 invece di 13 come risultato di $9 + 4$. Se Gauss voleva fare un calcolo più complicato, come ad esempio $7 \cdot 7$, il calcolatore a orologio gli restituiva il resto che si ottiene dividendo 49 (ossia $7 \cdot 7$) per 12. Il risultato è di nuovo 1. Ma era quando Gauss voleva calcolare $7 \cdot 7 \cdot 7$ che la potenza e la velocità del calcolatore a orologio cominciavano a emergere. Invece di moltiplicare un'altra volta 49 per 7, Gauss poteva limitarsi a moltiplicare per 7 l'ultimo risultato ottenuto, cioè 1, per ottenere la risposta, cioè 7. Così, senza dover calcolare $7 \cdot 7 \cdot 7$ (che fa 343), egli sapeva con poca fatica che quel risultato diviso per 12 dava resto 7. Il calcolatore dimostrò tutta la sua potenza quando gauss cominciò a utilizzarlo con grandi numeri […]. Pur non avendo idea di quanto facesse 7^{99}, il suo calcolatore a orologio gli diceva che quel numero diviso per 12 avrebbe dato resto 7.
> Gauss si rese conto che non c'era nulla di speciale negli orologi con 12 ore sul quadrante. Perciò introdusse l'idea di un'aritmetica dell'orologio (o aritmetica modulare, come viene a volta chiamata) basata su orologi con un numero qualsiasi di ore.>>

6. I NUMERI PRIMI

In questa sezione ci occuperemo dei numeri primi. Inizieremo illustrando il loro ruolo all'interno della TDN per poi analizzare risultati particolari su di essi dai quali si possono trarre dei procedimenti per stabilire la *primalità* di un intero $n > 1$ assegnato, ovvero la sua proprietà di essere primo.
Rimandiamo chi è interessato a questi argomenti al libro di Ribenboim ([17]) che fornisce una panoramica completa ed esauriente su tutte le questioni concernenti i primi, oltre che la primalità e la fattorizzazione.

6.1 LA SEQUENZA DEI PRIMI E LA FUNZIONE π

6.1.1 Numeri primi – analisi qualitativa

Nella sezione precedente si erano definiti i numeri primi nel modo che segue.

> "Un intero $n \geq 2$ si dice primo se possiede solo divisori banali, cioè se è divisibile solamente per 1 e n."

Ci si può logicamente chiedere come mai una definizione così semplice che esprime una caratteristica particolare di un intero – cioè l'essere *primo* – abbia tormentato molte tra le più grandi menti matematiche per parecchi secoli. Una risposta ovvia è quella che i numeri primi sono i "più piccoli" dopo 1 nella relazione di divisibilità: sono quelli che hanno meno divisori degli altri.
Osservando la successione dei numeri primi all'interno di quella degli interi positivi

$$1, \mathbf{2}, \mathbf{3}, 4, \mathbf{5}, 6, \mathbf{7}, 8, 9, 10, \mathbf{11}, 12, \mathbf{13}, 14, 15, 16, \mathbf{17}, 18, \mathbf{19}, 20, \ldots$$

o anche separatamente

$$2,3,5,7,11,13,17,19,23,29,31,37,41,43,47, \ldots$$

a parte l'apparente illogicità non si pensa al perché di tanto interesse. In fondo, che differenza può mai esserci tra, ad esempio, il numero 17 e il numero 18?
Nessuna, almeno fino a quando non si inizia a tracciare un quadro più ampio della situazione.

Rivediamo brevemente il teorema fondamentale dell'aritmetica (§5.1.2):

> "Qualunque naturale $n \geq 2$ si decompone in uno e un solo modo – a meno dell'ordine dei fattori – come prodotto di numeri primi."

In quest'ottica, il numero 18 si decompone come $18 = 2 \cdot 3^2$ mentre $17 = 17$ e basta. Tenendo fede al teorema, ci si accorge ben presto che i numeri primi sono gli atomi che danno origine agli altri numeri e che ogni intero positivo si esprime nel prodotto di fattori primi.
Tuttavia questo interesse per i primi può ancora intendersi come puramente *artistico* o come la mania dei matematici di cercare un ordine o una legge per qualsiasi cosa. In realtà i primi hanno molte proprietà particolari che fuoriescono dalla TDN e, tra l'altro, fanno molto comodo alla crittografia. Il sistema comunemente utilizzato per la protezione dei dati, infatti, è detto RSA – acronimo di Rivest, Shamir e Adleman, i 3 matematici che l'hanno creato – e si basa proprio su alcune proprietà tra congruenze e numeri primi. Per ora, quindi, diamo per scontato l'esistenza di un interesse più concreto sulla comprensione dei numeri primi.
Ma andiamo con ordine e ripartiamo dagli interi. A partire da un qualsiasi intero (positivo) n, ci si può porre le seguenti domande che danno origine ad altrettante problematiche.

- n è primo?
- Qual è la decomposizione di n in fattori primi?

A prima vista possono sembrare domande uguali o, quantomeno simili, tuttavia le differenze non tardano ad arrivare:

- se sappiamo la decomposizione di n in fattori primi, sappiamo anche se n è primo o no proprio perché se n è primo, la sua scomposizione è un banalissimo $n = n$ e viceversa;
- se sappiamo che n è primo o no, la scomposizione...?

Quest'ultima domanda resta in sospeso. Se ho una risposta alla domanda "n è primo o no" non è detto che in automatico conosco la sua decomposizione in fattori primi. E' ovvio che se n è primo, so anche che la sua decomposizione è $n = n$ ma se n è composto?
Vediamo nuovamente il caso di $n = 18$.

- Dire $18 = 2 \cdot 3^2$ ci fa capire anche che non è primo perché la sua scomposizione non è quella banale.
- Se invece so che 18 non è primo non risolvo nulla per la scomposizione e occorre fattorizzarlo a parte (con l'algoritmo che si preferisce).

Il problema della scomposizione di n – detta anche fattorizzazione – non è altro che quello di trovare una procedura che consente di ottenere la sua decomposizione in fattori primi ed è ben diverso dal problema della primalità che non è altro che il rispondere alla domanda "n è primo?".
Ci saranno molti risultati che rimarcheranno la differenza tra fattorizzazione e primalità. Vedremo, infatti, che la maggior parte delle procedure per la verifica della primalità di un numero sono completamente estranee alla fattorizzazione.

6.1.2 Una legge per i numeri primi

La prima domanda, quella più spontanea quando ci si trova faccia a faccia con i numeri primi, è se questa classe di numeri abbia una logica o, in altri termini, se ci sia una legge matematica che ne rappresenti la successione.
La risposta è stata oscura per molto tempo. Certo, l'interesse verso successioni che potevano dare numeri primi – come quella dei numeri di Mersenne o di Fermat (che vedremo nei prossimi paragrafi) – non si era mai spento. Tuttavia una risposta definitiva in questo ambito venne dal lavoro di un gruppo di matematici: Jones, Sato, Vada e Wiens che nel 1976 riuscirono a trovare un polinomio di grado 25 a 26 variabili i cui valori positivi erano tutti e soli i numeri primi ([8], §8; [13], §1.1.5).
Ovviamente era una formula piuttosto teorica e difficilmente applicabile nella realtà e si cercò di semplificarla senza molto successo. Infatti:

- aumentando il numero di variabili si diminuiva il grado del polinomio (con 45 variabili si ottiene una formula di quinto grado trovata dai matematici già citati in precedenza);
- diminuendo il numero di variabili si aumentava di conseguenza il grado del polinomio (J. Matijasievič portò le variabili a 10 ma il grado del polinomio ottenuto era circa $1{,}6 \cdot 10^{45}$).

Tale legge, dunque, esiste, ma è utile solo dal punto di vista teorico.
Tuttavia, come detto, esistono dei particolari numeri primi che destano più interesse rispetto agli altri e successioni particolari di numeri primi. Nel corso dei secoli, accanto all'interesse generale per lo studio dei numeri primi, si affiancava un interesse a delle successioni che assumevano – o *sembravano* assumere – valori primi al variare degli interi nel dominio.
Nei prossimi paragrafi analizzeremo le più famose in tal senso.

6.1.3 Numeri di Fermat

Nell'ottica appena accennata, una tra le successioni più celebri e studiate è senz'altro quella dei numeri di Fermat. Dato n, l'n-esimo numero di Fermat è definito nel modo seguente:

$$F_n = 2^{2n} + 1.$$

Possiamo notare che, per i primi valori di questa successione:

- $F_0 = 2^{2^0} + 1 = 2 + 1 = 3$, numero primo;
- $F_1 = 2^{2^1} + 1 = 4 + 1 = 5$, numero primo;
- $F_2 = 2^{2^2} + 1 = 16 + 1 = 17$, numero primo;
- $F_3 = 2^{2^3} + 1 = 2^8 + 1 = 256 + 1 = 257$, numero primo;
- $F_4 = 2^{2^4} + 1 = 2^{16} + 1 = 65536 + 1 = 65537$, numero primo.

A questo punto, Fermat ipotizzò, sulla base di questi, che la successione desse sempre valori primi. Tuttavia venne ben presto smentito da Eulero che, calcolando F_5 trovò la sorpresa:

$$F_5 = 2^{2^5} + 1 = 2^{32} + 1 = 4294967296 + 1 = 4294967297 = 641 \cdot 6700417.$$

Al giorno d'oggi gli unici primi di Fermat conosciuti sono proprio F_0, F_1, F_2, F_3 ed F_4, anche se si è arrivati a controllare valori molto avanzati della successione come $F_{2543548}$ nel Giugno 2011 ad opera di J. Scott Brown: questo numero che possiede 765687 cifre ([19]).

Teorema (Eulero)
Ogni numero di Fermat F_n non primo ha un divisore del tipo $k \cdot 2^{n+1} + 1$.

Questo risultato velocizza notevolmente la ricerca di un divisore per F_n sapendo che la successione dei numeri di Fermat cresce in maniera spropositata (F_5 ha 10 cifre, F_{10} ne ha più di 300!) e diventa ben presto ardua anche per una rete di calcolatori moderni.
E' stato grazie a questo teorema che lo stesso Eulero provò che

$$F_5 = 2^{2^5} + 1 = 2^{32} + 1 = 4294967296 + 1 = 4294967297 = 641 \cdot 6700417$$

smentendo clamorosamente la congettura di Fermat che tutti gli F_n fossero numeri primi.
In questa fattorizzazione, infatti, si può notare che

$$641 = 640 + 1 = 10 \cdot 2^6 + 1.$$

Teorema (Lucas) ([11], §2.6)
Ogni numero di FermatF_n non primo ha un divisore del tipo $k' \cdot 2^{n+2} + 1$.

Questo risultato è un perfezionamento del precedente poiché Lucas si rese conto che il fattore k menzionato da Eulero doveva essere pari, dunque ulteriormente divisibile per 2. Si può notare che questa nuova stima non è in contrasto con quanto detto nell'esempio precedente:

$$641 = 640 + 1 = 10 \cdot 2^6 + 1 = 5 \cdot 2^7 + 1.$$

Tuttavia sussistono altri risultati non meno interessanti circa la successione dei numeri di Fermat:

- $(F_n, F_m) = 1, \forall n, m \in \mathbb{N}, n \neq m$ (Goldbach);
- $\forall n \in \mathbb{N}, F_n$ è primo se e solo se F_n divide $3^{\frac{F_n - 1}{2}} + 1$ (Pepin).

In conclusione, non si sa ancora molto riguardo ai numeri di Fermat. Gli unici numeri di Fermat che sono anche numeri primi sono gli stessi che già noti a Fermat e cioè F_n per $n \leq 4$. Tuttavia non è escluso che ce ne siano altri: ci si chiede infatti quanti siano i primi (o i composti) di Fermat e se ce ne siano o meno di infiniti ([17], §2.6).
C'è dell'ottimismo per gli sviluppi futuri in questo senso. Per esempio un risultato di Sierpinski (1958) mostra come se un numero del tipo

$$S_n = n^n + 1, \qquad \forall n \in \mathbb{N}$$

è primo, allora $n = 2^{2^m}$ e cioè S_n è un numero di Fermat.
Come sempre la dura realtà non tarda a smentire l'entusiasmo poiché anche n^n cresce molto rapidamente al crescere di $n \in \mathbb{N}$. Inoltre, per ora, non si sa nemmeno se esistano altri numeri primi della forma S_n al di fuoridei primi ai quali corrispondono altrettanti primi di Fermat ([17], §2.6).

6.1.4 Numeri di Mersenne

Probabilmente, il caso di Mersenne è uno di quelli in cui Matematica e Musica si incontrano ([8], §2). Ogni nota, infatti, ha una certa frequenza e chi ha una certa dimestichezza con lo studio di uno strumento sa che il La sopra al Do di centro – quello del *diapason* tanto per capire – ha una frequenza di $440Hz$. Raddoppiando o dimezzando questa frequenza si ottiene la stessa nota ma, rispettivamente, un'ottava sopra o sotto.
Qualcuno potrebbe notare che questo modo di intendere le frequenze e le note musicali non è molto diverso, almeno in apparenza, a ciò che si è introdotto parlando di aritmetica modulare. In realtà, però, così non è in quanto la differenza tra una nota e quella immediatamente successiva (il semitono sopra) non si ottiene come una somma, come nei resti, ma con una moltiplicazione.
Vediamo di chiarire brevemente quest'ultima affermazione con un esempio, fissando $n = 12$. La scelta $n = 12$ non è casuale poiché chiunque abbia familiarità con la musica sa che un'ottava è composta da 12 semitoni. Se affermiamo, per $m \in \mathbb{N}$

$$m \equiv 7 \pmod{12}$$

vuol dire che dividendo m per 12 ottengo resto 7 ed era proprio l'idea alla base della classe resto. La scrittura 7_{12}, infatti, era l'insieme di tutti quei naturali che, divisi per 12, danno proprio resto 7 (§5.2.2)

$$7_{12} = \{m \in \mathbb{Z}, m \equiv 7 \pmod{12}\}.$$

La classe resto immediatamente successiva è quella dei numeri congrui a 8 modulo 12 e la si può intendere come quella attuale a cui si aggiunge un'unità. Nelle note, invece, se ho un Sol, potrei tranquillamente affermare – con un ragionamento simile a quello delle congruenze – che tutti i Sol sono equivalenti (a parte l'ottava!). Tuttavia, il semitono successivo, cioè il Sol#, si ottiene moltiplicando la frequenza del Sol per una costante ($\sqrt[12]{2}$) e non sommandola un qualsiasi tipo di unità come accade nell'aritmetica dell'orologio.
Per chi fosse interessato ad approfondire la questione appena accennata rimandiamo alla lettura di un qualsiasi articolo (wikipedia va benissimo) che parli di "Temperamento Equabile".
Nel sedicesimo secolo, cioè all'epoca di Mersenne, non si parlava ancora di Temperamento poiché questo sarebbe stato un argomento caldo nel secolo successivo anche grazie al lavoro di Bach. Tuttavia l'arte musicale era stata abbondantemente affinata nel corso dei secoli e molte idee erano già le stesse di oggi. Probabilmente, Mersenne ([8], §2) osservò che gli accordi "più dissonanti" erano quelli che si ottenevano dalle ottave perfette – che a questo punto possiamo pensarle come potenze del 2 – aggiungendo o sottraendo un semitono.
Nonostante la differenza con l'aritmetica modulare, egli pensò di portare questa semplice osservazione alla matematica – magari intendendo i numeri primi come un disordine, cioè una dissonanza rispetto all'ordine – occupandosi di studiare una sequenza del tipo

$$M_m = 2^m - 1, \qquad m \in \mathbb{N}.$$

Mersenne si accorse subito che questa sequenza non poteva produrre numeri primi se m era composto. Dal calcolo polinomiale imparato alle scuole superiori, infatti, sappiamo che

$$x^m - y^m = (x - y)r(x, y),$$

cioè la differenza di potenze è divisibile per la differenza della base. Però, nel caso di m composto

$$x^m - y^m = (x^a)^b - (y^a)^b = (x^a - y^a) \cdot r(x, y),$$

nella quale si è inteso $m = a \cdot b$ con $a, b > 1$ per definizione di m composto. Applicando questo ragionamento alla sequenza di Mersenne sapendo che $1 = 1^m, \forall m \in \mathbb{N}$ otteniamo

$$2^m - 1 = 2^m - 1^m = (2-1) \cdot r = r,$$

nel caso di m primo mentre

$$2^m - 1 = 2^m - 1^m = (2^a)^b - (1^a)^b = (2^a - 1^a) \cdot r = (2^a - 1) \cdot r,$$

nel quale $2^a - 1 > 1$, poiché $a > 1$.

Nel secondo caso, dunque, $2^m - 1$ è divisibile per un fattore intero maggiore compreso tra 2 e $m - 1$, quindi è composto.

Tuttavia, imporre m primo nella sequenza M_m non garantisce M_m primo. Per $m = 11$, infatti

$$M_{11} = 2^{11} - 1 = 2048 - 1 = 2047 = 23 \cdot 89.$$

A differenza dei numeri di Fermat, la successione di Mersenne produce altri numeri primi al crescere dei valori di n. Lo stesso Mersenne, nel 1640, asserì che M_m era primo per $m = 2,3,7,13,17,19,31,67,127,257$. Come altre frasi di grandi Matematici anche questa continua a provocare molto stupore:

$$M_{257} = 2^{257} - 1 \cong 2{,}31584 \cdot 10^{77}.$$

Ci si chiedeva, infatti, come avesse fatto con i mezzi dell'epoca a stabilire che un numero di 77 cifre era primo. Tuttavia si scoprì che per $m = 67, 257$, M_m non è primo mentre Mersenne dimenticò di citare i casi "primi" $m = 61,89,107$.

Analogamente ai numeri di Fermat, anche per quelli di Mersenne ci sono dei risultati interessanti trovati nel corso dei secoli che ne semplificano lo studio e l'eventuale scomposizione ([11], §2.6; [17], §2.7).

- Se m è primo e $m \equiv 3 \pmod 4$, allora $2m + 1$ divide M_m se e solo se $2m + 1$ è primo (Eulero-Lagrange).
- Se q divide M_m allora $q \equiv \pm 1 \pmod 8$ e $q \equiv 1 \pmod m$.

I più grandi attualmente conosciuti sono $M_{24036583}$ e $M_{25964951}$, quest'ultimo ha più di 7 milioni di cifre!

6.1.5 Numeri perfetti

Consideriamo $n \geq 1$ intero.

n è detto perfetto se è uguale alla somma di tutti i suoi divisori d, con $d < n$. Se tale somma è inferiore a n il numero è chiamato difettivo, altrimenti è detto sovrabbondante.

Vediamo di fare qualche esempio.

- 6 è un numero perfetto. $6 = 1 + 2 + 3$ nel quale 1,2,3 sono i divisori propri di 6.
- 12 è un numero sovrabbondante. $12 < 1 + 2 + 3 + 4 + 6 = 16$ nel quale 1,2,3,4,6 sono i divisori propri di 12.
- 14 è un numero difettivo. $14 > 1 + 2 + 7 = 10$ nel quale 1,2,7 sono i divisori propri di 14.

Gli unici numeri perfetti minori di 10000 sono 6,28,496,8128. Per ora si conoscono solo numeri perfetti pari ed è ancora un mistero se esistano o meno numeri perfetti dispari.

Si è scoperto che i numeri perfetti sono strettamente legati a quelli di Mersenne: vale, infatti, il seguente teorema.

Teorema ([11], §2.6)
Per ogni intero positivo m, $M_m = 2^m - 1$ è primo se e solo se $2^{m-1} \cdot M_m$ è un numero perfetto pari.

In questo modo studiare i numeri perfetti (pari) o la primalità dei numeri di Mersenne è la stessa cosa in quanto il primo ambito è strettamente collegato all'altro e viceversa.
Il numero 6, ad esempio, lo si può intendere come

$$6 = 2 \cdot 3 = 2^{2-1} \cdot M_2,$$

stessa cosa per il 28

$$28 = 4 \cdot 7 = 2^{3-1} \cdot M_3.$$

Attualmente non si conoscono numeri perfetti dispari; in generale non ci sono nemmeno dei risultati teorici che li escludono. Si attendono sviluppi futuri.

6.1.6 Perché sempre le potenze del 2?

Ci si può chiedere perché destano tanto interesse le potenze del 2 anche nello studio di possibili sequenze che producono numeri primi.
Le potenze del 2 sono state sempre oggetto di studio. Esse offrono un gran numero di divisori fin da numeri piccoli ($16 = 2^4$) e si riescono a calcolare con relativa semplicità anche con carta e penna rispetto alle altre.
Tralasciando l'informatica che si basa sul codice binario, un motivo plausibile è che dato p primo, $p > 2$ è dispari e

$$p^k \pm 1, \qquad k \in \mathbb{N}$$

è pari, dunque composto!
Sequenze altrettanto studiate sono quelle del tipo $q \cdot 2^k \pm 1$ oltre al già citato caso di $n^n \pm 1$ per n pari, altrimenti anch'esso è divisibile per 2.
Dal lavoro di Fermat sono state create sequenze di numeri detti numeri di Fermat generalizzati definiti nel modo seguente ([17], §5.7):

$$a^{2n} + 1, \qquad a \geq 2 \text{ pari}, \qquad n \geq 1.$$

Al contrario di quelli di Fermat, hanno prodotto molti numeri primi anche grandi al variare di n e b. Si può notare che a pari equivale a dire $a = q \cdot 2^k$ per un k opportuno.

6.1.7 Numeri di Germain

Sophie Germain era una matematica francese contemporanea di Gauss.

Nel sedicesimo secolo, il matematico Fermat affermò che non esistono radici intere non banali per l'equazione

$$x^n + y^n = z^n, \qquad n \geq 3 \text{ intero},$$

detta proprio equazione di Fermat.
Di essa, infatti, sono evidenti soluzioni banali (quando ad esempio x è nullo e $y = z$), mentre per soluzione non banale sene intende una composta da interi x, y, z non nulli.
Questa affermazione è conosciuta come *Ultimo Teorema di Fermat*.
Per molti matematici successivi questo teorema ebbe il sapore della beffa poiché Fermat, sul bordo di una pagina di un libro di Diofanto che stava leggendo, affermò di "avere una dimostrazione meravigliosa di questo fatto ma di non poterla scrivere sul margine del foglio per mancanza di spazio".
Si è discusso molto sull'attendibilità di questa rivendicazione, fino a quando, nel 1994, Wiles riuscì a dimostrare finalmente quella congettura. Tuttavia molti tentativi parziali vennero fatti nel corso dei secoli dimostrando il teorema in casi particolari ($n = 3$ è dovuto a Eulero).
Anche Sophie Germain ebbe un ruolo nella ricerca di questa dimostrazione. Di lei si ricorda un risultato che esclude l'esistenza di radici intere non banali per l'equazione di Fermat

$$x^p + y^p = z^p,$$

per p primo tale che $2p + 1$ è ancora primo.
I primi p tale che $2p + 1$ sono ancora numeri primi sono detti primi di Germain e di essi non si sa ancora se siano finiti o infiniti.

6.1.8 Altre sequenze

In quest'ultimo paragrafo enunceremo alcuni risultati interessanti circa delle semplici sequenze che danno o, meglio, "possono dare" origine a numeri primi. Singolarmente, ci sono altre formule studiate in passato in ambito della TDN che non saranno trattate qui. Per chi fosse interessato, si rimanda alla lettura del libro di Ribenboim ([17]).
Verranno esposti alcuni risultati che, anche singolarmente, contribuiranno a dimostrare che esistono infiniti numeri primi e, dunque, il teorema di Euclide già visto nella sezione precedente.

Teorema ([10], §2.3)
Ci sono infiniti primi della forma $4n + 3$ con $n \in \mathbb{N}$.

Questo risultato, così come il prossimo, può essere inteso nel seguente modo: la successione $4n + 3$ produrrà infiniti numeri primi. Esso, così come i seguenti dimostra il teorema di Euclide poiché i primi della forma $4n + 3$ non sono altro che un sottoinsieme di tutti i possibili numeri primi. In molti altri testi, questo teorema (così come i seguenti) è enunciato nel modo seguente: esistono infiniti primi p tale che $p \equiv 3 \pmod 4$.
Un risultato che si può dimostrare anche con conoscenze di scuola superiore è che i numeri primi possono essere congrui solamente a ± 1 modulo 6.

Sebbene, però, ci siano molti risultati intermedi (per esempio il fatto che esistono infiniti primi della forma $8n + 5$, con $n \in \mathbb{N}$), in generale vale il seguente

Teorema (Dirichlet) ([10], §2.3)
Se a, b sono degli interi positivi primi tra loro, allora ci sono infiniti primi della forma $an + b$, con $n \in \mathbb{N}$.

Questo risultato è onnicomprensivo degli altri e dimostra che tutte le funzioni lineari di $n$con coefficienti primi tra loro danno infiniti numeri primi. Logicamente si deve assumere $(a, b) = 1$ poiché, altrimenti, ogni termine dalla successione $an + b$ sarebbe divisibile per (a, b).
Ovviamente, un teorema del genere dimostra anche che ci sono infiniti primi dispari – cioè della forma $2n + 1$, con $n \in \mathbb{N}$ – confermando così il teorema di infinità dei numeri primi (tenendo conto che 2 è l'unico primo pari).

6.1.9 Primi gemelli

Due primi gemelli sono due numeri primi la cui differenza vale 2, in altre parole sono quelle coppie $(p, p + 2)$ nei quali p e $p + 2$ sono entrambi primi.
Si può notare che, a parte la coppia $(2,3)$, la minima differenza tra due primi consecutivi è proprio 2 proprio perché tra di loro c'è un numero pari. E' un problema aperto la questione se queste coppie di numeri siano infinite o meno.
Coppie di primi gemelli sono $(3,5), (5,7)$ ma anche $(41,43)$, per esempio. Come già detto, la nomenclatura "gemelli" trae origine dal fatto che 2 è la distanza minima possibile – a parte la coppia $(2,3)$ – tra due differenti numeri primi.
In analogia ai numeri primi "classici", si è introdotta la seguente funzione ([17], §4.3):

$$\pi_2(x) = \text{numero dei } p \leq x \text{ primi t. c. } p + 2 \text{ è primo},$$

che è la funzione enumerativa dei primi gemelli.
Per essa vale la seguente stima di Brun:

$$\pi_2(x) < \frac{100x}{\log^2(x)}.$$

Una caratterizzazione alternativa dei primi gemelli è quella di Clement del seguente teorema.

Teorema (Clement) ([17], §4.3)
Sia $n \geq 2$ intero. Allora$n$ e $n + 2$ formano una coppia di primi gemelli se e solo se

$$4[(n-1)! + 1] + n \equiv 0 \ \big(\text{mod } n(n+2)\big).$$

Come molti altri risultati di questo tipo, si tratta più di una proprietà teorica che di un utilizzo pratico poiché il calcolo di $(n-1)!$ consta di $(n-1)$ moltiplicazioni che diventano impraticabili per n molto grande.

6.1.10 Primi cugini e sexy

Accanto ai primi gemelli esistono altri tipi particolari di primi studiati dai matematici.
Due numeri primi sono detti cugini se p e $p+4$ sono entrambi primi. In altre parole una coppia di numeri primi cugini è del tipo $(p, p+4)$ dove entrambi sono primi.
Il nome "cugini", in analogia al caso dei primi gemelli, deriva dal fatto che 4 è la minore distanza che separa due primi dopo il 2. Un esempio è la coppia $(7,11)$.
Per questioni di aritmetica modulare, non esistono terne di numeri primi del tipo $(p, p+2, p+4)$, o $(p, p+4, p+8)$ poiché uno dei numeri della terna deve per forza essere divisibile per 3.
Diremo che due numeri primi sono sexy se la loro differenza è 6.
In questo caso la nomenclatura non è data dalla distanza ma deriva dal latino "sex" che sta a significare "sei". Un esempio può essere la coppia $(5,11)$.
Al contrario dei numeri primi gemelli e cugini, possono esistere anche terne, quadruple e anche quintuple (in realtà una sola) di primi sexy.

- Le terne di primi sexy sono del tipo $(p, p+6, p+12)$ con tutti e 3 i numeri primi e $p+18$ composto. Un esempio è $(17,23,29)$.
- Le quadruple di primi sexy sono del tipo $(p, p+6, p+12, p+18)$ con tutti e quattro i numeri primi. Una quadruple di primi sexy deve per forza iniziare con un numero primo la cui cifra finale è 1, altrimenti ci sarebbe un multiplo del 5 all'interno della quaterna stessa ([18]). L'eccezione è la quadrupla $(5,11,17,23)$.

L'unica quintupla di primi sexy è $(5,11,17,23,29)$. Non ce ne possono essere altre poiché all'interno di esse ci sarebbe un numero divisibile per 5 (la cui cifra finale è il 5, per l'appunto).

6.1.11 La funzione $\pi(x)$

Introduciamo la funzione enumerativa dei primi $\pi(n)$ definita nel modo seguente per ogni intero n:

$$\pi(n) = \text{numero dei primi} \leq n.$$

Vediamo qualche esempio.

- $\pi(2) = 1$;
- $\pi(3) = 2$, poiché 2,3 sono i primi ≤ 3 (2 in tutto);
- $\pi(10) = 4$, poiché 2,3,5,7 sono i primi ≤ 10 (quindi 4 di numero).

Anche se i numeri primi sono interi positivi, la funzione π si estende senza problemi ai reali, basta porre $\pi(x) = \pi(\lfloor x \rfloor)$, per $x \in \mathbb{R}$ (eventualmente non intero), ricordando che con la scrittura $\lfloor x \rfloor$ intendiamo il più grande intero $\leq x$. In particolare $\pi(x) = 0, \forall x < 2$. La funzione così ottenuta è localmente costante e aumenta di uno ogniqualvolta che x, crescendo verso $+\infty$, incontra un nuovo numero primo.
Otteniamo, ad esempio:

- $\pi(4{,}1) = \pi(4) = 2$;
- $\pi(\sqrt{52}) = \pi(7{,}2111\ldots) = \pi(7) = 4$, poiché 2,3,5,7 sono i primi ≤ 7 (e sono 4).

Sapere quanto vale $\pi(n)$ per n intero positivo (e trovare una procedura per il calcolo di $\pi(n)$) risponde alla domanda di quanti siano i numeri primi $\leq n$ che, tra l'altro, è l'oggetto dell'articolo di Riemann ("Sul numero dei primi minori di una certa quantità data", come dice anche il nome) che analizzeremo nelle sezioni seguenti.
Ci si chiede, dunque, se esiste una formula esplicita per il calcolo di $\pi(n)$ o, comunque, un procedimento accessibile per calcolare $\pi(n)$.
Esiste una formula "teorica"

$$\pi(n) = \sum_{p \leq n, p \text{ primo}} 1,$$

che, però non serve a nulla e, anzi, complica ancora di più le cose passando per il riconoscimento dei primi tra 2 e n.
Nel corso dei secoli ci sono stati diversi tentativi di trovare una formula o un procedimento per il calcolo di $\pi(n)$ senza servirsi della sua definizione teorica.
Tuttavia, spostando il discorso dagli interi ai reali (estendendo π ai reali nel modo descritto), si sono trovate nel corso dei secoli approssimazioni per la π, in particolare per il suo comportamento asintotico quando x tende a $+\infty$. La più famosa appartiene a Gauss e, espressa in termini moderni afferma che

$$\lim_{x \to \infty} \frac{\pi(x)}{x/\log(x)} = 1.$$

Riportiamo, ora, a titolo di esempio, la seguente tabella.

x	$\pi(x)$	$x/\log(x)$	$\pi(x) \Big/ \frac{x}{\log(x)}$
10	4	4,3	0,93
10^2	25	21,7	1,15
10^3	168	144,9	1,16
10^4	1229	1086	1,11
10^5	9592	8686	1,10
10^6	78498	72464	1,08
10^7	664579	621118	1,07
10^8	5761455	5434780	1,06
10^9	50847534	48309180	1,05

In questa tabella sono riportati i dati relativi al numero dei primi minori di potenze del dieci e dell'approssimazione data dal teorema dei numeri primi. Si vede (4° colonna) come il rapporto tra $\pi(x)$ e la sua approssimazione sia via via sempre più allineato verso un'uniformità davvero sorprendente se si considera l'apparente estraneità tra i numeri primi e il logaritmo naturale.
La congettura di Gauss venne dimostrata, in seguito, indipendentemente da J. Hadamard e Ch. De La Vallée Poussin e, da allora, chiamata con il nome di Teorema dei Numeri Primi. Essa è molto più di un semplice andamento oggettivo o di un casuale incontro tra due concetti

apparentemente lontani come il numero dei primi $\leq x$ e il logaritmo naturale. Ne forniremo una dimostrazione nell'appendice III di questa tesi.
Tuttavia, altri matematici si sforzarono di fornire una stima più "reale", valida effettivamente per ogni$x$, ottenendo così vari risultati piuttosto notevoli. Chebyshev, ad esempio, dimostrò che ([11], §2.4) $\forall x \geq 3$,

$$\frac{x}{2\log(x)} \leq \pi(x) \leq \frac{2x}{\log(x)}.$$

Questo risultato merita molta più attenzione di quella che può sembrare. La stima asintotica

$$\lim_{x\to+\infty} \frac{\pi(x)}{x/\log(x)} = 1,$$

approssima $\pi(x)$ con $\frac{x}{\log(x)}$ senza però fornire un'indicazione attendibile sull'errore che ne deriva. La dimostrazione di Chebyshev, invece, ci dà a partire da x un intervallo entro il quale abbiamo l'esatto valore di $\pi(x)$.
Si potrebbe tuttavia obiettare che:

$$\frac{2x}{\log(x)} - \frac{x}{2\log(x)} = \frac{4x-x}{2\log(x)} = \frac{3x}{2\log(x)} > \frac{x}{\log(x)}.$$

In altre parole l'intervallo nel quale avere il valore esatto di $\pi(x)$ risulta addirittura maggiore della sua approssimazione asintotica: si potrebbe concludere che l'errore nell'approssimazione dovuta a Chebyshev poteva anche superare il 100%!
Tuttavia, nel 1892, Sylvester ottenne un intervallo molto più piccolo rispetto a quello di Chebyshev ([17], §4.1) dimostrando che per x sufficientemente grande

$$0{,}95695\,\frac{x}{\log(x)} \leq \pi(x) \leq 1{,}04423\,\frac{x}{\log(x)}.$$

Di pari passo alla determinazione di approssimazioni o di metodi per calcolare la funzione $\pi(x)$, crescevano studi sulle stime tra differenti valori della stessa. Ne citiamo alcuni ([17], §4.1).

- $\pi(xy) \geq \pi(x) + \pi(y), x \geq y \geq 2, x \geq 6$ (Ishikawa, 1934).
- $\pi(x+y) \leq \pi(x) + \frac{2y}{\log y}$, sempre con $x \geq y \geq 2, x \geq 6$(Vaughan, 1962).
- $\pi(p_{n+1}^2) - \pi(p_n^2) > 4, n \geq 2$, in cui p_n denota l'n-esimo numero primo (Brocard, 1904).

Dal punto di vista teorico questi risultati sono interessanti: quello di Brocard, ad esempio, ci dice che tra i quadrati di due primi successivi esistono altri 4 numeri primi. Nella pratica, però, sono piuttosto dispersivi in quanto le stime sono molto deboli e non aiutano un calcolo efficace.
Un risultato decisamente curioso è dovuto ai progressi nei calcoli. Finora, infatti, sappiamo che per $x > 11, \pi(x) < \frac{x}{\log(x)}$ ed è un mistero se ci saranno dei valori che invertiranno questa tendenza.

6.2 RISULTATI E ALGORITMI PER LA PRIMALITA'

6.2.1 Primi, algoritmi e complessità

Il concetto di numero primo e il teorema fondamentale dell'aritmetica suggeriscono la seguente coppia di problemi: determinare algoritmi che, dato un input comune n naturale ≥ 2,

- decidono se n è primo o composto,
- calcolano la decomposizione di N nei suoi fattori primi.

Nel primo caso si parla di problema della primalità, nel secondo di problema della fattorizzazione. L'uno e l'altro ammettono algoritmi di soluzione noti fin dall'antichità. Ma la moderna teoria informatica della complessità computazionale ha sottolineato di disporre di algoritmi non solo eleganti nella teoria, ma anche efficaci nelle applicazioni pratiche. A questo proposito si accettano in genere come efficienti quelle procedure che impiegano tempi di lavoro al massimo polinomiali rispetto alla lunghezza dell'input, si bollano invece come troppo costose quelle che richiedono tempi almeno esponenziali rispetto alla stessa lunghezza.
In questa ottica è bene precisare quale è la lunghezza $l(n)$ di un numero naturale n (rappresentato, come usualmente si fa, in base 10, oppure in base 2, o in ogni altra base ammissibile). Essa è ovviamente il numero delle cifre di cui n si compone rispetto a quella base. Si vede che essa coincide approssimativamente col logaritmo di n in quella base. Per l'esattezza, quando la base è 10,

$$l(n) = \lfloor \mathrm{Log}(n) \rfloor + 1,$$

dove "Log" denota il logaritmo in base 10. Formule analoghe si hanno rispetto alle altre basi. Si ricordi tuttavia che, se a e b sono due di queste basi, i logaritmi di n rispetto alle basi a e b differiscono per una costante indipendente da n (il logaritmo di a rispetto a b). Altrettanto vale per le lunghezze, così che la scelta della base diventa in ultima analisi ininfluente per i nostri propositi.
I passi di una computazione su un input composto da uno o più numeri naturali si fanno invece coincidere con le operazioni elementari delle cifre di cui questi numeri si compongono. Si vede allora che l'addizione ha costo al più lineare, ovvero polinomiale di grado 1, rispetto alla lunghezza del massimo addendo, mentre la moltiplicazione ha costo al più quadratico, ovvero polinomiale di grado 2, rispetto alla lunghezza del massimo fattore. La sottrazione ha lo stesso costo di un'addizione, e la divisione lo stesso di una moltiplicazione. Costo quadratico ha anche la ricerca di massimo comune divisore e minimo comune multiplo (con l'algoritmo euclideo delle divisioni successive).

6.2.2 Un algoritmo elementare

Una prima procedura per stabilire la primalità di un numero è piuttosto intuitiva ed elementare e si può riformulare come segue. Sia $n > 1$ un intero.

- Si divide n per 2. Se la divisione viene precisa, cioè con resto 0, si deduce che n è composto, anzi divisibile per 2.
- Altrimenti si divide n per 3. Se la divisione dà resto zero si ha che n è composto e, anzi, abbiamo un suo fattore primo che è 3.
- Altrimenti si va avanti a dividere n per 4 con analoghe conclusioni.
- Iterando la procedura, si divide n per 5, e così via fino a che qualche divisore non dà resto nullo o si raggiunge $\sqrt{n}$.
- Se n non è divisibile per nessun intero positivo $d \leq \sqrt{n}$, allora n è primo, altrimenti n è composto.

L'algoritmo può indurre a trarre 2 conclusioni, entrambe sbagliate.

- Primalità e fattorizzazione sono la stessa cosa.
- L'una e l'altra si controllano in tempi ragionevoli, basta arrivare a un numero di operazioni $d \leq \sqrt{n}$ nel caso peggiore.

Per quanto riguarda il primo punto vedremo che ci sono algoritmi in grado di stabilire la primalità di un intero senza passare per la sua fattorizzazione.

Quanto al secondo osserviamo anzitutto che il controllo si può fermare a $\sqrt{n}$ perché, se un numero composto n si scrive come prodotto di due fattori positivi più piccoli, allora almeno uno dei due deve essere $\leq \sqrt{n}$. D'altra parte un controllo esteso a tutti i numeri interi compresi tra 2 e $\sqrt{n}$ richiede un numero di divisioni che, in linea di principio, è uguale a $\sqrt{n} - 1$ ed è dunque esponenziale rispetto alla lunghezza dell'input n (si ricordi che $\sqrt{n} = n^{1/2}$).

E' vero che un'ulteriore semplificazione è concessa: infatti, se si è già provato che n non è divisibile per un certo d, inutile controllare la sua divisibilità per i multipli di d. Ad esempio, se già sappiamo che n non è pari, è superfluo chiedersi se è multiplo di 4,6,8,…. Ma neppure queste facilitazioni riescono a ridurre significativamente i tempi di lavori (al variare di n per n grande).

6.2.3 Il crivello di Eratostene

Nel terzo secolo a.C. il matematico Eratostene – basandosi sull'idea (oggi largamente superata) che moltiplicare fosse meglio che dividere – ideò una semplice procedura atta a stabilire la primalità di un numero naturale.

Per illustrare il metodo di Eratostene consideriamo per esempio $n = 103$.

Il primo passo dell'algoritmo consiste nell'elencare in una tabella tutti i naturali dal 2 fino al numero di cui ci interessava sapere la primalità, nel nostro caso il 103 (in grassetto nella tabella).

	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100
101	102	**103**							

In seguito si operava escludendo tutti i multipli dei primi $\leq \lfloor\sqrt{n}\rfloor$ (nel nostro caso $\leq \lfloor\sqrt{103}\rfloor = \lfloor 10{,}148\ldots\rfloor = 10$) in ordine di grandezza dal primo più piccolo fino al più grande. Si iniziava, dunque, escludendo i multipli di 2, cioè i pari.

	2	3	~~4~~	5	~~6~~	7	~~8~~	9	~~10~~
11	~~12~~	13	~~14~~	15	~~16~~	17	~~18~~	19	~~20~~
21	~~22~~	23	~~24~~	25	~~26~~	27	~~28~~	29	~~30~~
31	~~32~~	33	~~34~~	35	~~36~~	37	~~38~~	39	~~40~~
41	~~42~~	43	~~44~~	45	~~46~~	47	~~48~~	49	~~50~~
51	~~52~~	53	~~54~~	55	~~56~~	57	~~58~~	59	~~60~~
61	~~62~~	63	~~64~~	65	~~66~~	67	~~68~~	69	~~70~~
71	~~72~~	73	~~74~~	75	~~76~~	77	~~78~~	79	~~80~~
81	~~82~~	83	~~84~~	85	~~86~~	87	~~88~~	89	~~90~~
91	~~92~~	93	~~94~~	95	~~96~~	97	~~98~~	99	~~100~~
101	~~102~~	**103**							

Si considerano poi i multipli del 3, poi del 5 e infine del 7, eliminandoli tutti, salvo ovviamente 3, 5, 7 e via dicendo. Nel nostro caso si termina escludendo i multipli di 7 poiché questo è il primo più grande $\leq \lfloor\sqrt{103}\rfloor = 10$.
Il risultato ottenuto è il seguente.

	2	3	~~4~~	5	~~6~~	7	~~8~~	~~9~~	~~10~~
11	~~12~~	13	~~14~~	~~15~~	~~16~~	17	~~18~~	19	~~20~~
~~21~~	~~22~~	23	~~24~~	~~25~~	~~26~~	~~27~~	~~28~~	29	~~30~~
31	~~32~~	~~33~~	~~34~~	~~35~~	~~36~~	37	~~38~~	~~39~~	~~40~~
41	~~42~~	43	~~44~~	~~45~~	~~46~~	47	~~48~~	~~49~~	~~50~~
~~51~~	~~52~~	53	~~54~~	~~55~~	~~56~~	~~57~~	~~58~~	59	~~60~~
61	~~62~~	~~63~~	~~64~~	~~65~~	~~66~~	67	~~68~~	~~69~~	~~70~~
71	~~72~~	73	~~74~~	~~75~~	~~76~~	~~77~~	~~78~~	79	~~80~~
~~81~~	~~82~~	83	~~84~~	~~85~~	~~86~~	~~87~~	~~88~~	89	~~90~~
~~91~~	~~92~~	~~93~~	~~94~~	~~95~~	~~96~~	97	~~98~~	~~99~~	~~100~~
101	~~102~~	**103**							

Poiché il 103 non è stato escluso, possiamo concludere che esso è primo. Non solo, in questa tabella tutti i numeri non esclusi sono primi: abbiamo, cioè, ottenuto una lista dei primi ≤ 103.

Generalizzare questo algoritmo è molto semplice. Supponiamo di avere $n \geq 2$.

- Si elencano tutti i numeri da 2 a n.
- Si escludono, da questa tabella, di volta in volta tutti i multipli dei primi compresi tra 2 e $\lfloor \sqrt{n} \rfloor$ in ordine crescente. I primi, dunque, sopravvivono e restano isolati.

Il risultato, oltre a dirci se n è primo o meno, elenca tutti i primi compresi tra 2 e n. La semplicità di questo algoritmo, tuttavia, è solo apparente. Esso poteva essere valido nel terzo secolo d.C. ma diventa inutile al giorno d'oggi. Infatti, per n grande – diciamo più grande di 10^{100} – il semplice elenco di tutti i naturali compresi tra 2 e n richiede tempi (e anche spazi) spropositati (di nuovo esponenziali rispetto alla lunghezza di n, quindi eccessivi). A proposito dello spazio necessario per scriverli, diciamo solo che è superiore alla memoria di un qualsiasi computer: anche se si limitasse a 1 Byte per numero (in media è molto di più) arriverebbe a 10^{100} Byte, cioè circa 10^{91} Gigabyte.

6.2.4 Il piccolo teorema di Fermat

Una proprietà notevole dei numeri primi è affermata dal così detto piccolo teorema di Fermat.

Teorema (piccolo teorema di Fermat) ([10], §6.1; [11], §3.5; [17], §2.2)
Sia p un numero primo. Allora per ogni intero a, vale

$$a^p \equiv a \pmod{p}.$$

In particolare, se p non divide a, allora $a^{p-1} \equiv 1 \pmod{p}$.

Il piccolo teorema di Fermat potrebbe allora ispirare il seguente algoritmo di primalità: dato un intero $p > 2$, si fissa casualmente un intero a compreso tra 1 e $p-1$ e si calcola innanzitutto (a, p).

- Se questo si prova diverso da 1, ci rivela che p è composto (e, tra l'altro, ci fornisce un suo divisore proprio).
- Se invece è 1, allora si procede a controllare $a^{p-1} \equiv 1 \pmod{p}$; se la congruenza non vale allora p è certamente composto. Ma cosa si può dedurre se la congruenza è soddisfatta?

Fu Sarrus a mostrare a questo proposito un risultato sconfortante:

$$2^{340} \equiv 1 \pmod{341},$$

ma $341 = 11 \cdot 31$ non è primo!

In questo caso si dice che 340 è uno pseudoprimo in base 2. In generale n intero dispari ($n > 2$) è uno pseudoprimo in base 2 se vale il piccolo teorema di Fermat, cioè

$$2^{n-1} \equiv 1 \pmod{n},$$

ma n non è primo.

Possiamo estendere la definizione appena vista. Se n è composto e per a intero primo con n vale $a^{n-1} \equiv 1 \pmod{n}$, diremo che n è uno pseudoprimo in base a. In altre parole n soddisfa la condizione del piccolo teorema di Fermat avendo a come base, ma n non è primo.

6.2.5 Gli pseudoprimi di Charmichael

Si può modificare il precedente algoritmo, facendo riferimento a <u>tutti</u> gli interi a primi con p (e compresi tra 1 e $p-1$). Per esempio, nel caso di 341 per il quale 2 non si dimostra un testimone attendibile, se invece si usa il 3 si ha

$$3^{340} \not\equiv 1 \pmod{341}.$$

Un altro esempio è $3^{90} \equiv 1 \pmod{91}$ mentre $2^{90} \not\equiv 1 \pmod{91}$, nei quali $91 = 7 \cdot 13$ non è affatto primo. E' lecito, dunque, porre la seguente domanda.

> <<Dato n intero $(n > 2)$ dispari, se per ogni a intero tale che $(a, n) = 1$ vale $a^{n-1} \equiv 1 \pmod{n}$ è giusto concludere n primo?>>

La risposta fu nuovamente negativa. Esistono dei numeri – detti pseudoprimi di Charmichael – che sono composti ma tali che

$$a^{n-1} \equiv 1 \pmod{n}, \qquad \forall a \text{ intero tale che } (a, n) = 1.$$

In altre parole gli pseudoprimi di Charmichael sono quei numeri composti per i quali vale la proprietà espressa nel piccolo teorema di Fermat a prescindere dalla base che si sceglie, purché quest'ultima sia prima con il numero considerato.
I più piccoli pseudoprimi di Charmichael sono 561, 1105 e 1729.

- 561 non è affatto primo. Scomposizione a parte, la somma delle cifre è un multiplo di 3 $(5 + 6 + 1 = 12)$ e quindi 561 è divisibile per 3.
- 1105 è divisibile per 5 avendo 5 come ultima cifra.
- 1729 non è primo, anche se ad occhio si poteva anche cadere nella trappola dato che non è divisibile né per 3, né per 5, né per 11.

Attorno al numero 1729, c'è un piccolo aneddoto che lo rende ancora più particolare ([8], §6). Si dice che quando Hardy andò a trovare Ramanujan malato all'ospedale, non sapendo come confortarlo, citò il numero del taxi che aveva preso, cioè 1729, come un numero privo di qualsiasi interesse. Tuttavia l'altro lo ammonì a non disprezzarlo poiché quello era il più piccolo numero esprimibile come somma di due diverse coppie di cubi. Infatti

$$1729 = 1 + 1728 = 1^3 + 12^3,$$
$$1729 = 1000 + 729 = 10^3 + 9^3,$$

il che dimostra come 1729 sia un numero notevole anche a prescindere dagli pseudoprimi di Charmichael.
Chi è interessato agli pseudoprimi di Charmichael troverà in ([11], §4.5) vari teoremi a loro proposito: si prova per esempio che essi sono infiniti e se ne danno delle caratterizzazioni puntuali.

Il piccolo teorema di Fermat ispira tuttavia vari algoritmi "completi" per il controllo della primalità. Alcuni sono troppo dispendiosi e quindi non molto utili dal punto di vista pratico. Per chi è interessato consigliamo di consultare ([11], §4.3) o ([17], §2.3).

6.2.6 La funzione ϕ di Eulero

Si è già discusso della funzione $\pi(x)$ e delle sue proprietà. Un'altra funzione importante nella teoria dei numeri è la funzione ϕ definita nel modo seguente: per ogni intero positivo n,

$$\phi(n) = \text{numero degli interi positivi} \leq n, \text{primi con } n.$$

Essa è chiamata la funzione ϕ di Eulero e, diversamente dalla già considerata $\pi(x)$, essa non si può estendere in modo naturale ai reali poiché riguarda strettamente la relazione di divisibilità tra interi.

Vediamo, ora, alcune sue importanti proprietà ([11], §3.6).

(i) $\phi(1) = 1$.

(ii) $\phi(p) = p - 1$, se p è primo. Infatti ogni intero compreso tra 1 e $p - 1$ è primo con p: da notare che se $p > 2$ è primo, $\phi(p)$ è un numero pari.

(iii) $\phi(p^k) = p^{k-1}(p - 1)$, per p primo e $k \geq 1$. Una dimostrazione semplice di questo fatto sta in [2] a pag. 61.

(iv) $\phi(n \cdot m) = \phi(n) \cdot \phi(m)$, per m, n interi primi tra loro.

Le proprietà (iii) e (iv) ci offrono, se combinate insieme, un algoritmo per il calcolo dei valori di $\phi(n)$, per n intero positivo. Sappiamo infatti che ogni numero n intero si può scrivere come prodotto di potenze di fattori primi a 2 a 2 distinti

$$n = \prod_{i=1}^{r} p_i^{a_i},$$

nel quale p_i sono i fattori primi che compongono n e a_i i corrispettivi esponenti. Possiamo, dunque, concludere

$$\phi(n) = \prod_{i=1}^{r} \phi\left(p_i^{a_i}\right) = \prod_{i=1}^{r} p_i^{a_i - 1}(p_i - 1).$$

Tuttavia questo metodo, per quanto valido dal punto di vista teorico, non è attuabile in pratica poiché necessita della conoscenza della scomposizione in fattori primi del numero n in questione.

Prima di andare avanti, possiamo fare qualche esempio.

(i) Per $n = 11$, abbiamo $\phi(11) = 10$: il numero 11 è primo, dunque $(a, 11) = 1$ per qualsiasi a compreso tra 1 e 10.

(ii) Se, invece, $n = 64$, essendo $64 = 2^6$, abbiamo $\phi(64) = 2^5 = 32$. Per una potenza del 2, anche il calcolo pratico è molto semplice: i numeri interi a compresi tra 1 e 63 e primi con il 64 sono tutti e soli i dispari. Tenuto conto che ad ogni pari segue un dispari e che gli estremi, cioè 1 e 63, sono entrambi dispari, abbiamo 31 numeri pari e 32 dispari. Dunque $\phi(64) = 32$.

(iii) Consideriamo $n = 28$, ricordando che $28 = 2^2 \cdot 7$,

$$\phi(28) = \phi(2^2) \cdot \phi(7) = 2 \cdot 6 = 12.$$

Per fare il raffronto pratico, sappiamo che i numeri interi compresi tra 1 e 27 e primi con 28 sono tutti tranne i multipli di 2 e 7, in altre parole i numeri dispari non divisibili per 7 e cioè 1,3,5,9,11,13,15,17,19,23,25,27 quindi 12 numeri.

Vediamo, di seguito, alcuni risultati importanti per la funzione ϕ.

Teorema
Per ogni intero positivo n, $n = \sum_{d|n} \phi(d)$.

Teorema (Eulero-Fermat) ([10], §6.1)
Se a e m sono interi positivi e $(a, m) = 1$, allora

$$a^{\phi(m)} \equiv 1 \pmod{m}.$$

Questo teorema è molto importante poiché è alla base del protocollo di crittografia RSA comunemente utilizzato in vari ambiti, in particolar modo nella sicurezza informatica.
Si noti che per m primo, $\phi(m) = m - 1$ quindi la nostra scrittura si riduce a

$$a^{m-1} \equiv 1 \pmod{m},$$

che non è altro che quella del piccolo teorema di Fermat vista in precedenza.

6.2.7 Altri teoremi sui primi e le congruenze

In questo paragrafo vedremo dei risultati importanti riguardanti i numeri primi e le congruenze. Tra essi, alcuni ispireranno intuitivamente degli algoritmi che si riveleranno inefficaci soprattutto perché non agevoli dal punto di vista della rapidità di esecuzione.

Teorema (Wilson) ([17], §2.2)
Un numero intero $p > 2$ è primo se e solo se

$$(p-1)! \equiv -1 \pmod{p},$$

ovvero, equivalentemente,

$$(p-1)! + 1 \equiv 0 \pmod{p}.$$

Questo risultato è importante perché conferma quanto già in parte intravisto con il teorema di Fermat e il successivo perfezionamento di Eulero: controllare la primalità di un intero e trovarne la fattorizzazione sono due problemi differenti.
La verifica della primalità con il teorema di Wilson richiede infatti di calcolare:

- un fattoriale $(p-1)!$
- la divisione implicita nella successiva congruenza

$$(p-1)! \equiv -1 \pmod{p}$$

anche se, procedendo in questo modo, nel caso che p si riveli composto non si ottiene informazione alcuna sulla sua decomposizione.

Tutto semplice, se non si considera il fatto che per calcolare $(p-1)!$ occorrono $p-2$ moltiplicazioni senza contare la forte crescita del fattoriale al crescere di n intero positivo. Così il criterio di Wilson è poco utile nella pratica, quand'anche sia ristretto alla verifica della sola primalità. Infatti $p-2$ è valore esponenziale rispetto alla lunghezza di p.
Si può osservare – il prossimo teorema ne è un altro esempio – che spesso risultati apparentemente semplici dai quali si può trarre un altrettanto semplice algoritmo di primalità, si rivelano essere inutili dal punto di vista pratico.

Teorema (Wolstenholme) ([10], §7.8)
Se $p \geq 5$ è primo, allora il numeratore di

$$1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{p-1}$$

è divisibile per p^2.

Sembrerebbe anche questo un algoritmo semplice da applicare per verificare la primalità di un numero: ma anche in questo caso ci si accorge ben presto della sua sostanziale inefficacia poiché, per stabilire l'eventuale primalità di p, occorre operare una somma di $p-2$ addendi, ovvero, nuovamente, un numero di operazioni esponenziale rispetto alla lunghezza dell'input.

Teorema
Se p è un numero primo, allora per ogni intero a compreso tra 1 e $p-1$

$$(x+a)^p \equiv x^p + a \pmod{p}.$$

Anche questo sembra, apparentemente, un risultato capace di originare semplici algoritmi di primalità. In questa espressione, infatti, x è una variabile (a valori interi), mentre a è un intero a scelta tra 2 e $p-1$. Il punto è che, sviluppare il binomio

$$(x+a)^p$$

richiede il calcolo di $p-1$ coefficienti per i termini intermedi tra x^p e a. Se, ad esempio, p fosse un numero dell'ordine di 10^{100}, occorrerebbe calcolare circa 10^{100} coefficienti e poi effettuare altrettante divisioni per p verificando per ciascuno di essi il sussistere della relazione di divisibilità da parte di p stesso. Se danno tutte esito positivo, quindi se l'equazione modulare è soddisfatta, allora vale p primo.

6.2.8 Equazioni con i moduli

In questo paragrafo introdurremo alcune problematiche relative ad equazioni con le congruenze.
Le equazioni che sono proposte tra le classi di resto modulo un dato intero positivo n e che di conseguenza sono dette "congruenziali" richiedono spesso una qualche attenzione. Hanno infatti tutte le specificità che derivano dal contesto cui si applicano. Consideriamo, per esempio, l'equazione

$$ax \equiv 1 \pmod{n}, \qquad a \in \mathbb{Z};$$

essa ci chiede di trovare l'inverso di a modulo n. Notiamo che tra gli interi a non ha inverso, a meno che a non coincida con $+1$ o -1.

Teorema
Se n è un numero primo, per $0 < a < n$, l'equazione

$$ax \equiv 1 \pmod{n}$$

ha sempre un'unica soluzione (modulo n).

Il teorema precedente si generalizza ad affermare che, per un intero $n > 1$, primo o composto, l'equazione $ax \equiv 1 \pmod{n}$ ha sempre un'unica soluzione (modulo n) purché a e n siano primi tra loro (condizione evidentemente soddisfatta quando n è primo e $1 < a < n$). Ancora più in generale si ha:

Teorema ([3], §5.3)
Se $(a, n) = 1$, allora l'equazione modulare

$$ax \equiv b \pmod{n},$$

ha un'unica soluzione.

La condizione $(a, n) = 1$ è fondamentale. Se, ad esempio $a = 2$ e $n = 4$, l'equazione

$$2x \equiv 3 \pmod{4},$$

non ammette soluzioni poiché per ogni intero x, $2x$ è un numero pari e non può essere congruo ad un dispari se n è pari.
Se già il problema è complicato per equazioni lineari, esso diventa molto più difficile nel caso in cui l'equazione modulare non è lineare. Consideriamo la seguente equazione di secondo grado:

$$x^2 \equiv 0 \pmod{n}.$$

Nel caso dell'uguaglianza non ci sarebbero problemi: $x = 0$ e siamo a posto.
Passiamo alla congruenza modulo n. Vediamo allora che anche $x = n$ è soluzione ($n^2 \equiv 0 \pmod{n}$); se, inoltre, n è un quadrato perfetto, anche $x = \sqrt{n}$ è soluzione. Se poi n è composto e col teorema fondamentale dell'aritmetica (§5.1.2) viene rappresentato come prodotto di potenze di primi distinti

$$n = \prod_{i=1}^{r} p_i^{a_i}, \qquad p_i \text{primo}, \qquad a_i \geq 1 \text{ intero},$$

allora si ottengono nuove soluzioni$x = a$ in cui

$$a = \prod_{i=1}^{r} p_i^{b_i}, \qquad p_i \text{ primo}, \qquad b_i \geq \left\lfloor \frac{1}{2} a_i + 1 \right\rfloor \text{intero}.$$

Se, infatti, andiamo a calcolare a^2, otteniamo

$$a^2 = \prod_{i=1}^{r} p_i^{2b_i} = k \cdot \prod_{i=1}^{r} p_i^{a_i} = k \cdot n, \qquad k \geq 1 \text{ intero},$$

proprio perché $2b_i \geq a_i$.
Vediamo di estendere il discorso all'equazione

$$x^2 \equiv a \pmod{n}, \qquad a < n.$$

Quanto detto per il caso $a = 0$ non si estende automaticamente ad altri interi a. Soffermiamoci nel caso particolare di

$$x^2 \equiv 1 \pmod{n},$$

che è abbastanza rappresentativo di quello che avviene in generale. La scelta di $a = 1$ è motivata dal fatto che ricorrerà spesso più avanti (per esempio, nell'ambito dei residui quadratici). In generale, per $a \neq 0$, la soluzione non è detto che ci sia – ne parleremo proprio trattando dei residui quadratici, però se a è un quadrato perfetto come quando $a = 1$, allora si hanno ovviamente le due soluzioni

$$x \equiv \pm 1 \pmod{n}.$$

Esse infatti valgono tra gli interi e di conseguenza tra gli interi modulo n. Per n primo possiamo dire di più.

Teorema

Dato n intero ($n \geq 2$), se n è primo, allora l'equazione

$$x^2 \equiv 1 \pmod{n}$$

ha esattamente 2 soluzioni, cioè $x \equiv \pm 1 \pmod{n}$.

Questo teorema, quindi, ci dice che se n è primo abbiamo 2 soluzioni, altrimenti ce ne potrebbero essere anche altre.

Vediamo di fare un paio di esempi.

Prendiamo $n = 7$, un numero primo. Il teorema appena visto ci assicura che l'equazione

$$x^2 \equiv 1 \pmod{7}$$

ha solo le due soluzioni $x \equiv \pm 1 \pmod{7}$: un esempio può essere il numero $36 = 6^2$

$$36 = 7 \cdot 5 + 1 \equiv 1 \pmod{7},$$

$$6 \equiv -1 \pmod{7}.$$

Sia ora $n = 8$, un numero composto. Ovviamente per l'equazione

$$x^2 \equiv 1 \pmod{8},$$

valgono lo stesso le due soluzioni $x \equiv \pm 1 \pmod{8}$: un esempio è $81 = 9^2$

$$81 = 10 \cdot 8 + 1 \equiv 1 \pmod{8},$$

$$9 \equiv 1 \pmod{8}.$$

Tuttavia vale anche $x \equiv 5 \pmod{8}$, infatti $5 \cdot 5 = 25 = 8 \cdot 3 + 1 \equiv 1 \pmod{8}$.

Quanto detto tornerà utile più avanti parlando dell'algoritmo probabilistico di Miller-Rabin per la ricerca della primalità di un numero. Infatti, scelto n intero di cui verificare la primalità, se si riesce a trovare una soluzione differente da $x \equiv \pm 1 \pmod{n}$ per l'equazione $x^2 \equiv 1 \pmod{n}$ si conclude automaticamente n composto.

6.2.9 Residui quadratici

Con il paragrafo precedente, abbiamo discusso in maniera non molto approfondita, ma esauriente di alcune problematiche riguardanti semplici equazioni modulari. Ora

descriveremo la teoria dei residui quadratici, che, a prescindere dal suo interesse specifico, è alla base dell'algoritmo probabilistico di primalità di Solovay-Strassen.
Sia n un intero positivo, in genere $n > 2$. Un intero a, primo con n si dice

- residuo quadratico modulo n se l'equazione $x^2 \equiv a \pmod{n}$ ammette soluzione, cioè se esiste un intero b tale che $b^2 \equiv a \pmod{n}$;
- non residuo quadratico modulo $n$altrimenti.

Possiamo osservare banalmente che 1 è sempre residuo quadratico poiché vale e $1^2 \equiv 1 \pmod{n}$ per qualsiasi n. Invece 0 è un quadrato modulo n per ogni n in quanto $0^2 \equiv 0 \pmod{n}$, ma 0 non è un residuo quadratico modulo n perché non è primo con n.
In generale, siccome si opera modulo n, possiamo restringere la nostra analisi al caso in cui a sia compreso tra 0 e $n-1$, altrimenti possiamo sostituirlo con il suo resto nella divisione per n. La terminologia "residuo quadratico" sottolinea questo punto, nel riferimento al "residuo" da intendere come "resto".
Consideriamo adesso il caso in cui $n = p$ è primo. Il caso $p = 2$ è semplice: tanto 0 quanto 1 sono quadrati modulo 2 e 1 è l'unico residuo quadratico modulo 2. Consideriamo allora p primo dispari.

Teorema ([11], §3.9)
Sia $p > 2$ un numero primo dispari. Allora tra le $p-1$ classi di resti modulo p di interi, $\frac{p-1}{2}$ corrispondono a residui quadratici modulo p e $\frac{p-1}{2}$ no.

In altre parole, tra tutte le classi resto modulo p, una metà sono residui quadratici e l'altra no.
Questi teoremi non sono definiti per $p = 2$ anche perché alcuni di loro sono banali ([10], §6.6). Nel caso del risultato appena enunciato, per $p = 2$ non è soddisfatto poiché sia 0 che 1 sono residui quadratici in quanto $0^2 = 0$ e $1^2 = 1$. Di questo fatto, inoltre, c'è una spiegazione molto più poetica: i quadrati dei numeri pari sono quelli congrui a 0 modulo 2 mentre quelli dei numeri dispari sono quelli congrui a 1 modulo 2.
Dato un primo dispari p, definiamo ora per ogni intero a il simbolo di Legendre $\left(\frac{a}{p}\right)$ di a rispetto a p nel modo seguente

- $\left(\frac{a}{p}\right) = 1$ se a è un residuo quadratico modulo p;
- $\left(\frac{a}{p}\right) = -1$ se a è un non residuo quadratico modulo p;
- $\left(\frac{a}{p}\right) = 0$ se a divide p.

Per il simbolo di Legendre valgono alcune proprietà importanti. Una delle quali è il seguente risultato di Eulero.

Teorema (Eulero) ([10], §6.6; [11], §3.9)
Sia p un primo dispari. Allora per ogni intero a,

$$\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \pmod{p}.$$

Tra l'altro, questo risultato conferma il piccolo teorema di Fermat. Infatti, ricordando i vari valori assunti dal simbolo di Legendre, si osserva che, se p non divide a,

$$a^{p-1} = a^{2\frac{p-1}{2}} \equiv 1 \pmod{p}.$$

Invece se p divide a, allora a ed ogni sua possibile potenza risultano congrui a 0 modulo p.
Viceversa, dalla precedente discussioni sull'equazione $x^2 \equiv 1 \pmod{p}$ segue che

$$a^{p-1} = a^{2\frac{p-1}{2}} \equiv 1 \pmod{p}$$

ha esattamente due soluzioni e cioè

$$a^{\frac{p-1}{2}} \equiv \pm 1 \pmod{p}$$

che, tra l'altro, sono i due valori assunti dal simbolo di Legendre quando a è primo con p.
La procedura derivante da questo teorema per il calcolo del simbolo di Legendre è rapida poiché si affida al calcolo delle potenze modulo p, che è appunto veloce (§5.2.3). Tuttavia nell'algoritmo di Solovay-Strassen si preferisce riservare il teorema di Eulero alla ricerca dei numeri primi (vedremo dopo il come e il perché) e affidarsi di conseguenza a procedure alternative per il calcolo preventivo del simbolo di Legendre. A questo proposito sono utili le seguenti osservazioni.

- Come già sottolineato, per ogni a, $\left(\frac{a}{p}\right) = \left(\frac{r}{p}\right)$, dove r è il resto della divisione di a per p.
- Per ogni scelta di interi a, b, $\left(\frac{a}{p}\right) \cdot \left(\frac{b}{p}\right) = \left(\frac{a \cdot b}{p}\right)$, in altre parole il simbolo di Legendre preserva il prodotto.

Queste osservazioni suggeriscono un algoritmo sufficientemente rapido per il calcolo del simbolo di Legendre:

- si passa da a al suo resto r nella divisione per n,
- si decompone poi r nel suo prodotto di potenze di fattori primi

$$r = \prod_{i=1}^{s} p_i^{a_i}, \qquad p_i \text{primo}, \qquad a_i \geq 1,$$

- si deduce

$$\left(\frac{a}{p}\right) = \left(\frac{r}{p}\right) = \prod_{i=1}^{s} \left(\frac{p_i}{p}\right)^{a_i}.$$

Torniamo al caso generale. Dopo i due passi già descritti, diventa fondamentale diventa fondamentale riuscire a calcolare il simbolo di Legendre nel caso di un numero primo a. Prima di procedere in questo premettiamo due semplici osservazioni.

- $\left(\frac{1}{p}\right) = 1$ perché 1 è sempre un residuo quadratico.
- $\left(-\frac{1}{p}\right) \equiv (-1)^{\frac{p-1}{2}} \pmod{p}$, come si prova direttamente o affidandosi al teorema di Eulero (basta considerare $a = -1$).

Passiamo allora, come promesso, al caso di a primo. Valgono qui due teoremi difficili e famosi (specie il secondo), i quali affermano quanto segue.

- $\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}$. Questo implica che 2 è un residuo quadratico quando $p \equiv \pm 1 \pmod 8$ altrimenti è un non residuo quadratico.
- Per p e q primi dispari distinti vale la legge della reciprocità quadratica di Gauss:

$$\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}\left(\frac{p}{q}\right).$$

Questo implica che il segno non cambia quando almeno uno dei due primi è congruo a 1 modulo 4. In generale si può notare che gli esponenti sono comunque interi poiché si è supposto p, q primi dispari.

Sembra allora che si sia ottenuto complessivamente l'algoritmo richiesto per il calcolo del simbolo di Legendre. Ma in verità il metodo appena delineato ha il difetto di affidarsi alla decomposizione in fattori primi di a e quindi si espone agli eventuali ritardi di questa fattorizzazione. C'è tuttavia una maniera di ovviare questo problema. Basta allargare gli orizzonti e considerare il simbolo di Jacobi che non è altro che un'estensione di quello di Legendre.
Lo introduciamo così: considerato $n \geq 2$ intero e a intero, poniamo

$$\left(\frac{a}{n}\right) = \prod_{i=1}^{r}\left(\frac{a}{p_i}\right)^{a_i},$$

scrittura che deriva dalla già citata rappresentazione di n come prodotto di fattori primi

$$n = \prod_{i=1}^{r}(p_i)^{a_i}, \qquad p_i \text{primo}, \qquad a_i > 0.$$

Possiamo notare che per n primo, il simbolo di Jacobi coincide con il simbolo di Legendre, tuttavia non ne preserva le proprietà e non si collega più alla proprietà di essere un quadrato modulo n. Per esempio se$n = 15$,

$$\left(\frac{2}{15}\right) = \left(\frac{2}{3}\right)\cdot\left(\frac{2}{5}\right) = -1 \cdot -1 = 1$$

ma 2, che pure è primo con 15, non è un quadrato modulo 15.
Valgono ancora tutte le proprietà già enunciate per il simbolo di Legendre che ricordiamo in breve.

(i) Per ogni a intero, $\left(\frac{a}{n}\right) = \left(\frac{r}{n}\right)$, dove$r$ è il resto della divisione di a per n.

(ii) Per ogni scelta di interi a, b, $\left(\frac{a}{n}\right)\cdot\left(\frac{b}{n}\right) = \left(\frac{a\cdot b}{n}\right)$.

(iii) $\left(\frac{2}{n}\right) = (-1)^{\frac{n^2-1}{8}}$, analogamente a quanto visto per il simbolo di Legendre.

(iv) Vale ancora la legge di reciprocità quadratica di Gauss per m e n interi dispari

$$\left(\frac{n}{m}\right) = (-1)^{\frac{n-1}{2}\cdot\frac{m-1}{2}}\left(\frac{m}{n}\right).$$

Per il calcolo del simbolo di Jacobi e, conseguentemente, del simbolo di Legendre, esiste una procedura agevole che fa leva sulle proprietà appena enunciate.
Se, infatti, dovessimo calcolare

$$\left(\frac{m}{n}\right), \qquad m, n \in \mathbb{Z}^+,$$

innanzitutto, qualora $m > n$, andremmo a sostituirlo con il suo resto della divisione con n, per la proprietà (i).
In seguito, inizieremo con l'estrarre – qualora fosse possibile – il fattore 2 (con la sua eventuale molteplicità) dall'intero m in base alle proprietà (ii) e (iii). Successivamente, tramite la (iv), potremo sempre invertire i fattori qualora non potessimo più andare avanti tenendo sempre a mente la proprietà (i) per sostituire il numeratore con l'eventuale resto della divisione con il denominatore.

6.2.10 L'algoritmo di Solovay-Strassen

L'algoritmo di Solovay-Strassen trae le sue fondamenta dal teorema di Eulero riferito al calcolo del simbolo di Legendre: se n è primo dispari, ogni intero a (in particolare ogni a primo con n) soddisfa la congruenza

$$a^{\frac{n-1}{2}} \equiv \left(\frac{a}{n}\right) (\mathrm{mod}\ n).$$

Non vale però il viceversa, in particolare se qualche intero a primo con n permette a n di soddisfare la precedente congruenza non possiamo con ciò dedurre che n è primo. Possono infatti esistere i così detti pseudoprimi di Eulero in base a, ovvero numeri n composti dispari che tuttavia rendono vera insieme ad a la congruenza di cui sopra.
Però, come nel caso del piccolo teorema di Fermat, ci si chiede che cosa succede facendo più prove, cioè ricorrendo a diversi valori di a. Nel caso del teorema di Fermat la risposta era negativa, ma stavolta le cose vanno meglio.
Vediamo, dunque, l'algoritmo un po' più nel dettaglio.

Algoritmo
Abbiamo n dispari, di cui vogliamo scoprire la possibile primalità. Il tutto si basa sulla formula del teorema di Eulero ricordando che il calcolo delle potenze e del simbolo di Jacobi sono procedure abbastanza rapide quando si ha a che fare con le congruenze.

- Si sceglie casualmente a compreso tra 1 e $n-1$.
- Se $(a,n) \neq 1$, si deduce che n è composto e l'algoritmo termina.
- Se $(a,n) = 1$ ma non vale $a^{\frac{n-1}{2}} \equiv \left(\frac{a}{n}\right) (\mathrm{mod}\ n)$ allora n è composto e l'algoritmo termina.
- Se $(a,n) = 1$ e $a^{\frac{n-1}{2}} \equiv \left(\frac{a}{n}\right) (\mathrm{mod}\ n)$, allora si dichiara che n è "<u>probabilmente</u> primo".

L'algoritmo è simile a quello che si serve del piccolo teorema di Fermat e la risposta in caso affermativo, anche qui, non è una certezza.
In quest'ottica, diremo che per n composto dispari e a intero, se vale

$$a^{\frac{n-1}{2}} \equiv \left(\frac{a}{n}\right) (\mathrm{mod}\ n),$$

n è uno pseudoprimo di Eulero in base a. Tuttavia, rispetto all'algoritmo che si è tentato di ricavare dal piccolo teorema di Fermat, per n composto le cose cambiano e non esistono equivalenti degli pseudoprimi di Carmichael.

Teorema ([11], §4.5)
Sia n intero positivi composto dispari. Allora per almeno metà degli $\phi(n)$ interi primi con n a due a due incongrui modulo n <u>non</u> vale la congruenza

$$a^{\frac{n-1}{2}} \equiv \left(\frac{a}{n}\right) \pmod{n}.$$

Quindi, se n è composto, scelto un intero a compreso tra 1 e $n-1$, la probabilità che non valga quella congruenza è al massimo $\frac{1}{2}$ e quindi l'eventualità di errore è al più $\frac{1}{2}$. Scelto un altro intero la probabilità di errore scende ad al più $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. In generale, dopo k tentativi, la possibilità di avere una risposta sbagliata – cioè dichiarare n primo quando invece non lo è – scende ad al più $\frac{1}{2^k}$.
Errore a parte, il teorema appena visto ci dice anche che se scegliamo più della metà degli interi compresi tra 1 e $n-1$ per verificare la primalità di n tramite l'algoritmo di Solovay-Strassen, abbiamo la certezza di avere una risposta esatta.
Può scandalizzare la natura probabilistica della procedura: l'affidarsi a testimoni a che possono anche mentire. Ma se l'algoritmo si sviluppa in modo rapido e il margine d'errore è tollerabile (100 tentativi lo riducono a un valore minore della probabilità di fare 5 volte 6 al superenalotto) allora la sua applicazione pratica sembra ragionevole.

6.2.11 Algoritmo di Miller-Rabin

L'algoritmo di Miller-Rabin, trae le sue fondamenta dall'analisi del piccolo teorema di Fermat combinato al fatto che l'equazione modulare $x^2 \equiv 1 \pmod{n}$ ha come uniche soluzioni $x \equiv \pm 1 \pmod{n}$ per n primo (§6.2.8).
Vediamolo con calma, per poi riassumerlo in modo schematico ([11], §4.6).
Siano n intero (dispari) di cui vogliamo verificare la primalità e $1 < a < n$ anch'esso intero. Ci interessa il caso $(a, n) = 1$ poiché, se non fosse $(a, n) = 1$ avremmo trovato un divisore comune tra a e n e, dunque, che n non è primo.
A questo punto verifichiamo la condizione del piccolo teorema di Fermat (§6.2.4)

$$a^{n-1} \equiv 1 \pmod{n},$$

anche in questo caso, se non è soddisfatta concludiamo che n è composto. Altrimenti possiamo andare oltre: se n è dispari, $n-1$ è pari, dunque

$$\left(a^{\frac{n-1}{2}}\right)^2 \equiv 1 \pmod{n}$$

cioè

$$a^{\frac{n-1}{2}} \equiv \pm 1 \pmod{n}.$$

Se poi $a^{\frac{n-1}{2}} \equiv 1 \pmod{n}$ e $\frac{n-1}{2}$ è anch'esso pari, otteniamo

$$\left(a^{\frac{n-1}{4}}\right)^2 \equiv 1 \pmod{n},$$

cioè

$$a^{\frac{n-1}{4}} \equiv \pm 1 \pmod{n}$$

e così via si itera la procedura fino a quando non si ottiene un risultato diverso da ± 1 oppure $\frac{n-1}{2^k}$ è dispari ($k \geq 1$ intero) e non si può andare oltre. Più precisamente, scriviamo

$$n - 1 = 2^k \cdot t$$

nel quale $k \geq 1$ è tale che 2^k è la maggiore potenza del 2 che si può estrarre da $n - 1$, in altre parole k è tale che $2^k | n$ ma $2^{k+1} \nmid n$. Allora t è, ovviamente, l'intero dispari che si ottiene dividendo $n - 1$ per 2^k. La relazione iniziale si traduce con

$$\left(a^{2^{k-1} \cdot t}\right)^2 \equiv a^{2^k \cdot t} \equiv 1 \pmod{n}$$

da cui deduciamo $a^{2^{k-1} \cdot t} \equiv \pm 1 \pmod{n}$. Se, inoltre, $a^{2^{k-1} \cdot t} \equiv 1 \pmod{n}$ e $k \geq 2$, allora si ha

$$\left(a^{2^{k-2} \cdot t}\right)^2 \equiv a^{2^{k-1} \cdot t} \equiv 1 \pmod{n}$$

e si itera il ragionamento. In generale, andando avanti l'algoritmo termina con una risposta positiva se

- incontriamo -1
- concludiamo con $a^t \equiv 1 \pmod{n}$

e con una risposta negativa altrimenti.
Ricapitoliamo i dettagli.

Algoritmo
Abbiamo $n > 2$ intero dispari. Estraendo la potenza massima del 2 da $n - 1$ poniamo $n - 1 = 2^k \cdot t$ con $k, t \geq 1$ naturali e t dispari. Scegliamo a intero, $1 < a < n$.

- Se $(a, n) \neq 1$, n è composto e l'algoritmo termina.
- Se $(a, n) = 1$ ma non vale $a^{n-1} \equiv 1 \pmod{n}$ allora n è composto, in analogia a quanto detto per l'algoritmo ricavato dal piccolo teorema di Fermat (§6.2.5).
- Se $(a, n) = 1$ e $a^{n-1} \equiv a^{2^k \cdot t} \equiv 1 \pmod{n}$ applichiamo le osservazioni fatte in precedenza. Se $a^t \not\equiv 1 \pmod{n}$ e $a^{2^r \cdot t} \not\equiv -1 \pmod{n}$ per ogni naturale $r \leq k$ allora n è composto.
- Se, invece, $(a, n) = 1$ e $a^{n-1} \equiv a^{2^k \cdot t} \equiv 1 \pmod{n}$ e, servendoci delle osservazioni precedenti troviamo $a^t \equiv 1 \pmod{n}$ oppure $a^{2^r \cdot t} \equiv -1 \pmod{n}$ per qualche $r \leq k$ dichiariamo che n è <u>probabilmente</u> primo.

Volendo, si può scegliere b intero, $1 < b < n$ e ricominciare l'algoritmo daccapo se si vuole ottenere una risposta più precisa.

Analizziamo l'output dell'algoritmo, traendo qualche conclusione significativa.
Un'eventuale risposta "n composto" è <u>sicura</u> perché deriva da violazioni di relazioni che risultano valide per n primo: $(a, n) = 1$ e le congruenze analizzate ad inizio paragrafo.

Se, invece, l'algoritmo termina con la risposta "n probabilmente primo", resta l'incertezza e l'eventualità che n sia composto. Analogamente ai casi precedenti, ci si chiede quanto è presente tale possibilità attuando l'algoritmo: purtroppo, anche in questo caso, esiste, cioè esistono numeri composti per i quali l'algoritmo dà come risultato "n probabilmente primo" anche se non è così.
Sia, dunque, n composto dispari > 2 tale che $n - 1 = 2^k \cdot t$ con $k, t \geq 1$ interi e t dispari e sia, inoltre, a un intero primo con n. Diremo che n è uno pseudoprimo forte in base a se $a^t \equiv 1 \pmod{n}$ oppure, per qualche naturale $r \leq k$ vale $a^{2^r \cdot t} \equiv -1 \pmod{n}$.
In altre parole uno pseudoprimo forte è proprio un numero composto per il quale l'algoritmo dà come risposta – sbagliando – n primo.
Purtroppo numeri del genere esistono anche se le cose sembrano andare decisamente meglio rispetto all'algoritmo di Solovay-Strassen visto nel paragrafo precedente. Tanto per cominciare, il più piccolo pseudoprimo forte nelle basi 2, 3, 5, 7 (contemporaneamente) è $n = 3215031751$.
Rispetto all'algoritmo precedente, quindi, la situazione migliora come i due seguenti risultati dimostrano.

Teorema
Se n è uno pseudoprimo forte in una qualche base a, con $(a, n) = 1$, allora n è anche uno pseudoprimo di Eulero in base a.

Questo teorema ci dice che qualora fallisse l'algoritmo di Miller-Rabin, allora anche quello di Solovay-Strassen darebbe una risposta sbagliata utilizzando la stessa base. Questo risultato, dunque, ci impedisce di applicare – magari in parallelo – entrambe le procedure servendosi dello stesso dato.

Teorema
La probabilità che n sia uno pseudoprimo forte in base a è per al più $\frac{1}{4}$ degli interi a primi con n compresi tra 1 e n.

Dunque la probabilità di errore che si ottiene con un singolo tentativo è al massimo di $\frac{1}{4}$. Iterando la procedura, dopo k tentativi, questa probabilità si riduce a $\frac{1}{4^k}$, la metà rispetto a quella di Solovay-Strassen. Inoltre, i calcoli richiesti dall'algoritmo di Miller-Rabin sono, come per Solovay-Strassen, ragionevolmente veloci.

6.2.12 Algoritmo AKS

L'algoritmo AKS – nome in codice per indicare i tre studiosi indiani Agrawal, Kayal e Saxena – è un algoritmo agevole e, soprattutto, deterministico: il risultato, infatti, non ha un margine di errore ma è certo.
Alla base di quest'algoritmo c'è la seguente identità (§6.2.7)

$$(x + a)^n \equiv (x^n + a) \pmod{n},$$

che vale per un generico a (intero) se e solo se n è un numero primo. In precedenza, però, si era visto che un algoritmo basato direttamente su una congruenza di questo tipo aveva scarso impatto nelle applicazioni pratiche.
La fortuna di questo nuovo algoritmo è di estendere questa idea a identità più semplici da controllare: in questo modo si accelera il processo riducendo la complessità (e la lunghezza) di questa equazione modulare avendo cura di preservare l'importante proprietà di base che possiede, cioè il fatto che vale se e solo se n è primo.
L'idea, dunque, è quella di lavorare non più *solo* modulo n, ma anche modulo $x^r - 1$ nel quale r è un numero ragionevolmente piccolo. Quindi, invece di calcolare $(x+a)^n$, si calcola questa potenza mediante il resto della divisione con $x^r - 1$.
La potenza, infatti, si esegue non nella sua totalità con tutti i coefficienti da calcolare, ma mediante l'algoritmo di Lagrange (§5.2.3) avendo cura di sostituire i risultati intermedi qualora essi siano di grado maggiore a $x^r - 1$, operando al riguardo la divisione tra polinomi nell'usuale modo che si impara dalle scuole secondarie superiori.
Il punto a favore di tutto il discorso è che il resto tra un qualsiasi polinomio di grado superiore a r e $x^r - 1$ (di grado r) ha un grado $< r$ e, dunque, un numero di termini $\leq r$ al contrario degli $n+1$ termini che si ottengono espandendo $(x+a)^n$ nella sua totalità.
Con una notazione largamente usata in questo caso, andremo ad esprimere quanto detto nel seguente modo:

$$(x+a)^n \equiv x^n + a \pmod{n, x^r - 1},$$

la quale significa proprio che invece dei termini originali, andiamo a confrontare i resti della divisione di entrambi con $x^r - 1$. Il prossimo teorema ci assicura che la condizione trovata possiede ancora la sua proprietà originale, cioè se continua a valere il fatto che è verificata se e solo se n è primo.

Teorema (Agrawal, Kayal, Saxena) ([11], §4.7)
Sia n un numero intero dispari $n > 2$. Sia poi un intero positivo $r < n$ tale che n ha periodo $> \log_2^2 n$ modulo r. Allora n è primo se e solo se valgono le seguenti condizioni.

(i) n non è una potenza perfetta.
(ii) n non ha fattori primi $\leq r$.
(iii) $(x+a)^n \equiv x^n + a \pmod{n, x^r - 1}$ per ogni intero positivo $a < \sqrt{n}\log_2 n$.

Prima di passare all'algoritmo occorre fare una precisazione. Dire che n ha un periodo $> \log_2^2 n$ modulo r equivale a dire che $n^k \equiv 1 \pmod{r}$ solo per $k > \log_2^2 n$.
Vediamo, dunque, i passi fondamentali di tale algoritmo.

Algoritmo di primalità AKS
Abbiamo in input un numero intero dispari $n > 2$.

- Se n è una potenza perfetta, cioè $n = a^s$ con $a, s \geq 2$ interi, dichiariamo n composto.
- Altrimenti troviamo il più piccolo r con periodo $> \log_2^2 n$ e calcoliamo (a, n) per tutti gli $a \leq r$. Se $(a, n) \neq 1$ per almeno un a, allora dichiariamo (ovviamente) n composto.

- In caso contrario, per $a = 1,2,\dots,\lfloor\sqrt{r}\log_2 n\rfloor$, verifichiamo la nostra congruenza $(x+a)^n \equiv x^n + a \pmod{n, x^r - 1}$: se per qualche a non è soddisfatta, allora dichiariamo n composto.
- Altrimenti dichiariamo n primo.

Il primo passo non è difficile, senza entrare nei dettagli diciamo che ci sono degli algoritmi rapidi che consentono di vedere se n è una potenza perfetta o meno.
Per il secondo passo, si prova che$r$ così scelto esiste e si può riconoscere in un numero di passi al più polinomiale rispetto alla lunghezza di n.
Il terzo passo richiede di verificare $\lfloor\sqrt{r}\log_2 n\rfloor$ congruenze che sono una quantità polinomiale rispetto alla lunghezza di n.
Senza entrare nel dettaglio, quindi, otteniamo una procedura che, nel complesso, lavora in tempo polinomiale rispetto alla lunghezza dell'input con cui si ha a che fare.
In generale, l'algoritmo deriva dal teorema di Agrawal, Kayal e Saxena e i suoi passi sono semplicemente legati alle implicazioni del teorema stesso: si cerca di violarle singolarmente, dichiarando n primo solo se risultano tutte soddisfatte. Per n composto, infatti, almeno una delle 3 condizioni (i), (ii), (iii) non sarebbe soddisfatta ([11], §4.7).
In questa sezione, l'algoritmo AKS è stato presentato a grandi linee, senza entrare nei dettagli tecnici che necessiterebbero, tra l'altro, di approfondire molti concetti che non sono trattati in questa tesi. L'obiettivo era quello di inquadrare i principali risultati in tema di verifica della primalità nel corso dei secoli per mostrare come quello della primalità sia un problema molto più complicato e vario di quello che si può credere ad una prima analisi.

7. COSTANTE DI EULERO-MASCHERONI

In questa breve sezione tratteremo di un costante molto particolare che occupa un posto tutto suo nella matematica come il numero π o il numero e. Essa è la costante di Eulero-Mascheroni e viene indicata con la lettera greca γ.

7.1 Esistenza della costante (γ)

Il nostro punto di partenza è la serie armonica semplice già discussa in precedenza nella sezione di Richiami di Analisi Matematica I (§1.2.4)

$$\sum_{k=1}^{\infty} \frac{1}{k} = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} + \cdots.$$

Di essa ci interessa analizzare la somma parziale n-esima:

$$S_n = \sum_{k=1}^{n} \frac{1}{k} = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}, \qquad n \in \mathbb{N}$$

e il suo legame con il $\log n$. Tale somma parziale, nel caso della serie armonica, è indicata anche con il simbolo H_n e chiamata "n-esimo numero armonico" ([24]). Dalla teoria delle serie esposta precedentemente sappiamo che

$$\lim_{n \to +\infty} S_n = \lim_{n \to \infty} \sum_{k=1}^{n} \frac{1}{k} = \sum_{k=1}^{\infty} \frac{1}{k}.$$

Ora mostreremo che la somma parziale n-esima della serie armonica è dello stesso ordine di $\log n$, più precisamente ([36])

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

la quale per $n \to +\infty$ si tradurrà con

$$\lim_{n \to +\infty} \left(\sum_{k=1}^{n} \frac{1}{k} - \log(n) \right) = \gamma.$$

La costante γ che ne deriva è detta appunto costante di Eulero-Mascheroni o, più semplicemente, anche numero di Eulero (anche se questa dicitura può far confondere con il numero di Nepero e).

Vediamo, dunque, di dimostrare quanto detto.

Innanzitutto fissiamo $n \in \mathbb{N}$ non nullo e riscriviamo $\log(n)$ nella seguente forma:

$$\log(n) = \sum_{k=1}^{n-1} (\log(k+1) - \log(k)).$$

Osserviamo, infatti, che i termini della somma al secondo membro si annullano a due a due al crescere degli indici. In altre parole nell'espressione

$$\sum_{k=1}^{n-1}(\log(k+1)-\log(k))$$
$$= (\log(2)-\log(1)) + (\log(3)-\log(2)) + (\log(4)-\log(3)) + \cdots$$
$$+ (\log(n)-\log(n-1))$$

il termine con argomento minore, cioè $\log 1$ è nullo per definizione di logaritmo, mentre tutti gli altri termini intermedi a segni alterni si eliminano a coppie fino a che resta $\log(n)$.
A questo punto ricordiamo la seguente proprietà del logaritmo:

$$\log\left(\frac{a}{b}\right) = \log(a) - \log(b), \qquad \forall a, b \in \mathbb{R}^+.$$

La somma al secondo membro dell'espressione precedente si traduce allora in

$$\sum_{k=1}^{n-1}(\log(k+1)-\log(k)) = \sum_{k=1}^{n-1}\log\left(\frac{k+1}{k}\right) = \sum_{k=1}^{n-1}\log\left(1+\frac{1}{k}\right).$$

Possiamo, dunque, considerare l'espressione da portare al limite sotto un'ottica differente

$$\lim_{n\to\infty}\left(\sum_{k=1}^{n}\frac{1}{k}-\log(n)\right) = \lim_{n\to\infty}\left(\sum_{k=1}^{n}\frac{1}{k}-\sum_{k=1}^{n-1}\log\left(1+\frac{1}{k}\right)\right) = \sum_{k=1}^{\infty}\left(\frac{1}{k}-\log\left(1+\frac{1}{k}\right)\right).$$

In essa l'ultimo passaggio è giustificato dal fatto che per $n \to +\infty$ anche $(n-1) \to +\infty$: questo ci consente di portare l'elemento all'interno della seconda sommatoria, nella prima.
Ora supponiamo di aver dimostrato che l'ultima espressione converge indicando con γ il suo limite e con γ_n la somma parziale n −esima. Allora, tenendo conto dell'uguaglianza

$$\sum_{k=1}^{n-1}\log\left(1+\frac{1}{k}\right) = \sum_{k=1}^{n}\log\left(1+\frac{1}{k}\right) - \log\left(1+\frac{1}{n}\right), \qquad \forall n \in \mathbb{N}\backslash\{0\},$$

abbiamo:

$$\sum_{k=1}^{n}\frac{1}{k}-\log(n) = \sum_{k=1}^{n}\frac{1}{k}-\sum_{k=1}^{n-1}\log\left(1+\frac{1}{k}\right) = \sum_{k=1}^{n}\left(\frac{1}{k}-\log\left(1+\frac{1}{k}\right)\right) + \log\left(1+\frac{1}{n}\right).$$

Se supponiamo di aver già dimostrato la convergenza, otteniamo i due seguenti risultati:

$$\sum_{k=1}^{n}\left(\frac{1}{k}-\log\left(1+\frac{1}{k}\right)\right) = \gamma_n,$$

per definizione di somma parziale n −esima e

$$\log\left(1+\frac{1}{n}\right) = o(1).$$

L'ultima vale per definizione di o piccolo (§1.3.3), infatti

$$\lim_{n\to+\infty}\frac{\log\left(1+\frac{1}{n}\right)}{1} = \lim_{n\to+\infty}\log\left(1+\frac{1}{n}\right) = \log(1) = 0.$$

Unendo le ultime considerazioni, otteniamo

$$\sum_{k=1}^{n}\frac{1}{k}-\log(n) = \gamma_n + o(1),$$

cioè

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

che è la seconda delle due formule che volevamo dimostrare. A questo punto occorre studiare la serie

$$\sum_{k=1}^{\infty} \left(\frac{1}{k} - \log\left(1 + \frac{1}{k}\right)\right),$$

per farlo ci serviremo della formula di Taylor applicata alla funzione $f(x) = \log(1 + x)$ per $x \in]0,1]$. Le seguenti osservazioni confermano che è lecito adoperarla.

(i) $\forall k \in \mathbb{N}\backslash\{0\}$, $1 + \frac{1}{k} \in]1,2]$, in particolare $\frac{1}{k} \in]0,1]$.

(ii) $\log\left(1 + \frac{1}{k}\right) \in]0, \log(2)] \subseteq]0,1], \forall k \in \mathbb{N}\backslash\{0\}$ per la (i).

(iii) La formula di Taylor applicata a $\log(1 + x)$ per $x \in]0,1]$ è valida anche per $\log\left(1 + \frac{1}{k}\right)$ per le osservazioni (i) e (ii). Infatti, $x = \frac{1}{k}$ è un caso particolare di $x \in]0,1]$ poiché i valori assunti dalla successione sono contenuti in $]0,1]$.

Scriviamo, dunque, la formula di Taylor centrata in 0 e troncata al primo ordine con il resto di Lagrange (§1.3.6)

$$\log(1 + x) = x - \frac{x^2}{2(1 + \eta)^2}, \qquad \eta \in]0, x], \qquad (x \in]0,1]).$$

Dunque

$$x - \log(1 + x) = \frac{x^2}{2(1 + \eta)^2}, \qquad \eta \in]0, x], \qquad (x \in]0,1]).$$

Siccome $\eta > 0$

$$2(1 + \eta)^2 > 2,$$

che si traduce nella seguente

$$x - \log(1 + x) = \frac{x^2}{2(1 + \eta)^2} < \frac{x^2}{2}, \qquad x \in]0,1].$$

Riprendendo la serie sotto esame, deduciamo

$$0 < \frac{1}{k} - \log\left(1 + \frac{1}{k}\right) < \frac{1}{2k^2}, \qquad k \geq 1, \qquad \left(\frac{1}{k} \in]0,1]\right).$$

Possiamo, quindi, maggiorare la serie di partenza termine a termine con il risultato appena ottenuto:

$$0 < \sum_{k=1}^{+\infty} \left(\frac{1}{k} - \log\left(1 + \frac{1}{k}\right)\right) < \sum_{k=1}^{+\infty} \frac{1}{2k^2} = \frac{1}{2} \sum_{k=1}^{+\infty} \frac{1}{k^2}.$$

L'ultima serie è convergente (§1.2.4). Quindi, grazie ai criteri di confronto per le serie (§1.2.4) concludiamo che anche la serie di partenza è convergente.

7.2 Osservazioni

La dimostrazione appena vista riguardo alla convergenza della serie

$$\sum_{k=1}^{+\infty}\left(\frac{1}{k}-\log\left(1+\frac{1}{k}\right)\right)$$

e della definizione della costante di Eulero-Mascheroni come limite dell'espressione stessa ci consente di fare qualche ulteriore osservazione.

- Il culmine della dimostrazione della convergenza della serie è l'espressione seguente:

$$0<\frac{1}{k}-\log\left(1+\frac{1}{k}\right)<\frac{1}{2k^2}, \qquad \forall k\geq 1.$$

Questa, applicata alla serie nella sua interezza, era il tassello finale che dimostrava la convergenza di tale serie, cioè

$$0<\sum_{k=1}^{+\infty}\left(\frac{1}{k}-\log\left(1+\frac{1}{k}\right)\right)<\sum_{k=1}^{+\infty}\frac{1}{2k^2}=\frac{1}{2}\sum_{k=1}^{+\infty}\frac{1}{k^2}.$$

L'ultima maggiorazione ci fornisce una prima stima della costante di Eulero-Mascheroni:

$$0<\gamma<\frac{1}{2}\sum_{k=1}^{+\infty}\frac{1}{k^2}=\frac{\pi^2}{12}\cong 0{,}822467.$$

Tuttavia l'unica utilità di questa approssimazione è l'individuazione di un primo limite superiore per la γ. Ricordiamo che l'ultimo risultato che abbiamo utilizzato, cioè il valore esatto della serie $\sum_{k=1}^{+\infty}\frac{1}{k^2}$, è un famoso teorema dovuto anch'esso al matematico Eulero. (Vedi, ad esempio, ([1])).

- Per definizione di somma parziale n-esima applicata alla serie possiamo valutare la differenza tra γ e γ_n nel modo che segue.

$$\begin{aligned}0\leq\gamma-\gamma_n&=\sum_{k=1}^{\infty}\left(\frac{1}{k}-\log\left(1+\frac{1}{k}\right)\right)-\sum_{k=1}^{n}\left(\frac{1}{k}-\log\left(1+\frac{1}{k}\right)\right)\\&=\sum_{k=n+1}^{\infty}\left(\frac{1}{k}-\log\left(1+\frac{1}{k}\right)\right)<\sum_{k=n+1}^{\infty}\frac{1}{2k^2}=\lim_{m\to\infty}\sum_{k=n+1}^{m}\frac{1}{2k^2}\\&<\lim_{m\to\infty}\sum_{k=n+1}^{m}\int_{k-1}^{k}\frac{dx}{2x^2}=\lim_{m\to\infty}\int_{n}^{m}\frac{dx}{2x^2}=\lim_{m\to\infty}\left(-\frac{1}{2m}+\frac{1}{2n}\right)=\frac{1}{2n}.\end{aligned}$$

In essa si è usata la seguente maggiorazione per giungere all'integrale finale:

$$\int_{k-1}^{k}\frac{dx}{2x^2}=-\frac{1}{2x}\bigg|_{k-1}^{k}=-\frac{1}{2k}+\frac{1}{2(k-1)}=\frac{-2(k-1)+2k}{2k(k-1)}=\frac{1}{2k^2-2}>\frac{1}{2k^2},$$

che vale $\forall k>1$.

Infine ricordiamo che, per una qualsiasi funzione (integrabile) $f(x)$

$$\begin{aligned}\sum_{k=n+1}^{m}\int_{k-1}^{k}f(x)dx&=\int_{n}^{n+1}f(x)dx+\int_{n+1}^{n+2}f(x)dx+\cdots+\int_{m-1}^{m}f(x)dx\\&=\int_{n}^{m}f(x)dx,\end{aligned}$$

per l'additività dell'integrale.

7.3 Conclusione

La costante di Eulero-Mascheroni è stata definita come limite di una serie. Il suo valore numerico, troncato alle prime venti cifre decimali è

$$\gamma \cong 0{,}57721566490153286060651 20 \ .$$

Di per sé non si sa ancora se sia razionale o meno. Ammette comunque, insieme alla definizione appena data come limite di una serie, altre rappresentazioni, spesso tramite integrali ([14]).

Per ciò che riguarda questa tesi, la costante di Eulero-Mascheroni ricorrerà nelle sezioni future a proposito delle funzioni Gamma di Eulero e Zeta di Riemann.

8. LA FUNZIONE GAMMA

In questa sezione tratteremo una funzione particolare che troverà ampio spazio all'interno dell'ipotesi di Riemann. Si tratta della funzione Gamma, definita dal matematico Eulero.
La trattazione di questa funzione meriterebbe uno spazio molto più approfondito di quello offerto in queste pagine, dove ci limiteremo alla definizione e alla comprensione delle proprietà principali della stessa per non divagare eccessivamente dagli intenti di questa tesi.
In quest'ottica sarà omessa la maggior parte delle dimostrazioni: alcune perché complicate, altre perché eccessivamente lunghe e "non interessanti" dal punto di vista della tesi stessa.

8.1 Introduzione

Consideriamo, dato un numero $n \in \mathbb{N}$, il fattoriale $n!$ definito per n positivo come

$$n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1 = \prod_{k=1}^{n} k,$$

e, per $n = 0$, come $0! = 1$.
La definizione del fattoriale è molto semplice e immediata, tuttavia il calcolo pratico di quest'ultimo è una vera spina nel fianco per i calcolatori poiché l'unica formula "agevole" per ottenerlo è proprio quella data dalla definizione. Questo implica che per calcolare, ad esempio, $400!$ occorre applicare la formula ottenendo

$$400! = 400 \cdot 399 \cdot 398 \cdot \ldots \cdot 3 \cdot 2 \cdot 1 \cong 6{,}4 \cdot 10^{868},$$

che è un numero di 868 cifre difficile anche da immaginare.
Tuttavia per i moderni calcolatori l'ostacolo maggiore non è tanto la grandezza del numero ottenuto quanto la lentezza dell'algoritmo che si usa per calcolarlo. Il calcolo di $n!$, infatti, implica $n-1$ moltiplicazioni, ovvero un valore spropositato per n grande. Certo, anche la grandezza del numero è un problema: i moderni calcolatori consentono di computare $n!$ in maniera agevole anche per $n \leq 10000$, ma poi la rapida crescita del fattoriale (per esempio, $10001! = 10001 \cdot 10000!$ ha 5 cifre in più di $10000!$) crea ben presto dei problemi.
Nel Settecento si iniziò a trattare più a fondo il problema e la domanda a cui dare una risposta divenne, ben presto, la seguente ([3]):

"$\exists f: I \subseteq \mathbb{R} \to \mathbb{R}$ t.c. $f(n) = n!, \forall n \in \mathbb{N}$?"

In altre parole ci si chiedeva se esistesse una funzione di variabile reale che, ristretta ai naturali, desse come risultato il fattoriale. La risposta era senz'altro affermativa, però il problema si complicava se l'oggetto della ricerca era una funzione piuttosto regolare, cioè per lo meno di classe C^1. Una soluzione positiva di questa domanda poteva aprire nuovi orizzonti

nel calcolo stesso di $n!$ avendo a disposizione una legge che avrebbe potuto agevolarlo evitando di passare per il prodotto n-esimo.
La soluzione venne da più parti (Legendre, Eulero, Weierstrass...) e condusse all'individuazione della funzione Γ (Gamma).

8.2 Definizione (in $\mathbb{R}$) e proprietà

La funzione Γ è quella che associa ad ogni reale positivo x il seguente valore ([19], §8.17)

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt.$$

Questa definizione è ben posta poiché l'integrale converge uniformemente per $x > 0$ ([12]; [5], §6.1).
Vediamo, ora, alcune importanti proprietà.

(i) $\Gamma(1) = 1$.
Andiamo a calcolarlo a partire dalla definizione della Γ.

$$\Gamma(1) = \int_0^\infty t^{1-1} e^{-t} dt = \int_0^\infty e^{-t} dt = \lim_{b\to\infty} \int_0^b e^{-t} dt = \lim_{b\to\infty} -e^{-t}|_0^b$$
$$= \lim_{b\to\infty} (-e^{-b} + 1) = 1.$$

(ii) Per ogni reale $x > 1$, $\Gamma(x) = (x-1)\Gamma(x-1)$.
Per dimostrarlo basta effettuare un'integrazione per parti nella definizione stessa della Gamma:

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt = -t^{x-1} e^{-t}|_0^\infty - \int_0^\infty (-(x-1) t^{x-2} e^{-t}) dt$$
$$= (x-1) \int_0^\infty t^{x-2} e^{-t} dt = (x-1)\Gamma(x-1).$$

(iii) Dalla (ii) $\Gamma(n) = (n-1)\Gamma(n-1)$, $\forall n \in \mathbb{N}, n \geq 2$. Unendola alla (i) abbiamo che la funzione Gamma è una generalizzazione del fattoriale naturale. Infatti, per n intero positivo, si ha

$$\Gamma(n) = (n-1)\Gamma(n-1) = (n-1)(n-2)\Gamma(n-2) = \cdots = (n-1)!.$$

Attenzione, però, a non confondere gli indici: non vale $\Gamma(n) = n!$ ma $\Gamma(n) = (n-1)!$.

(iv) Per $x > 0$, vale la seguente formula di duplicazione trovata dal matematico Legendre

$$\Gamma(2x) = \frac{2^{2x-1}}{\sqrt{\pi}} \Gamma(x) \Gamma\left(x + \frac{1}{2}\right),$$

la cui dimostrazione è omessa poiché si serve di strumenti più avanzati rispetto a quelli richiamati in questa sezione.

(v) Per $x, y > 0$, allora

$$\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} = \int_0^1 t^{x-1} (1-t)^{y-1} dt.$$

Una dimostrazione di questa identità la si può trovare in molti testi, ma una delle più semplici sta in ([5], §6.1).
Generalmente, si pone

$$\mathrm{B}(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)},$$

nel quale la funzione $\mathrm{B}(x,y)$ è detta "funzione Beta di Eulero" ([28], §A.3).

(vi) Se nella precedente uguaglianza poniamo $x = y = \frac{1}{2}$, otteniamo

$$\frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma(1)} = \Gamma\left(\frac{1}{2}\right)^2 = \int_0^1 t^{\frac{1}{2}-1}(1-t)^{\frac{1}{2}-1}dt = \int_0^1 t^{-\frac{1}{2}}(1-t)^{-\frac{1}{2}}dt.$$

Operiamo, ora, la sostituzione $t = \sin^2(x)$, da cui $dt = 2\sin(x)\cos(x)$:

$$\int_0^1 t^{-\frac{1}{2}}(1-t)^{-\frac{1}{2}}dt = \int_0^{\frac{\pi}{2}} (\sin^2(x))^{-\frac{1}{2}}(1-\sin^2(x))^{-\frac{1}{2}}(2\sin(x)\cos(x))dx$$

$$= \int_0^{\frac{\pi}{2}} \frac{2\sin(x)\cos(x)}{\sin x}(\cos^2(x))^{-\frac{1}{2}}dx = \int_0^{\frac{\pi}{2}} \frac{2\sin x\cos x}{\sin x\cos x}dx$$

$$= \int_0^{\frac{\pi}{2}} 2dx = 2x\Big|_0^{\frac{\pi}{2}} = \pi,$$

e deduciamo $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

8.3 Estensioni della funzione Gamma al piano complesso ($z \neq 0$)

La funzione Gamma può essere estesa in modo naturale al semipiano complesso:

$$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt\,, \qquad z \in \mathbb{C}\ \mathrm{t.c.}\ Re(z) > 0.$$

Questa estensione è semplice poiché si ottiene semplicemente considerando z complesso al posto di x reale: inoltre per $z = x \in \mathbb{R}$ ci dà esattamente la funzione Γ introdotta nel paragrafo precedente.

Tuttavia, per l'unicità del prolungamento analitico (§3.2.6) essa è l'unica estensione della funzione Gamma al semipiano complesso. In questo modo, tutte le proprietà esposte in nella sezione precedente continuano a valere anche nel caso $z \in \mathbb{C}$, tale che $Re(z) > 0$.

Tuttavia il matematico Eulero riuscì ad estenderla analiticamente all'intero piano complesso. Il ragionamento di Eulero fu quello di considerare la definizione di funzione Gamma proprio come estensione del fattoriale naturale. Il risultato fu il seguente:

$$\Gamma(z) = \frac{\Gamma(z+n)}{z\cdot(z+1)\cdot\ldots\cdot(z+n-1)}, \qquad n \in \mathbb{N} \text{ tale che } Re(z+n) > 0$$

e vale per ogni $z \in \mathbb{C}\backslash\{0\}$ a patto di trovare un naturale n che soddisfi insieme a z l'ultima condizione.

Questa idea è semplice e immediata e mostra come la funzione Gamma, estesa al piano complesso, ha dei poli semplici in corrispondenza degli interi non positivi, cioè $z = 0, -1, \ldots$.

In questi punti, il residuo della funzione Gamma è il seguente:

$$Res(\Gamma, -n) = \frac{(-1)^n}{n!}.$$

Analogamente a Eulero, anche Gauss riuscì a trovare una rappresentazione della Γ come limite di un prodotto

$$\Gamma(z) = \lim_{n\to\infty} \frac{n!\, n^z}{z(z+1)\cdot\ldots\cdot(z+n)}.$$

A partire da quest'ultima, si riesce a trovare la seguente sotto forma di prodotto infinito ([3])

$$\Gamma(z) = \frac{1}{z}\prod_{n=1}^{\infty} \frac{(1+1/n)^z}{1+z/n},$$

ed è ricordato anche per il seguente risultato.

Teorema (Principio di Riflessione di Eulero)
Se $z \in \mathbb{C}$, allora

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)}.$$

Questa formula fornisce un'estensione della Γ all'intero piano complesso (tranne $z \neq 0$) in un modo molto più rapido rispetto alla prima idea di Eulero. Una definizione alternativa è dovuta a Weierstrass e si basa su una rappresentazione mediante un prodotto infinito ([1], §5.2.4):

$$\Gamma(z) = \frac{e^{-\gamma z}}{z}\prod_{n=1}^{\infty}\left(1+\frac{z}{n}\right)^{-1} e^{\frac{z}{n}},$$

nella quale $\gamma \cong 0{,}577216$ è la costante di Eulero-Mascheroni (§7).
Questa rappresentazione, inoltre, è l'unica che mostra esplicitamente che la funzione Gamma non ha degli zeri: ricordiamo, infatti, che $a^z \neq 0, \forall a, z \in \mathbb{C}$, $(a \neq 0)$.

Teorema (formula di Stirling generalizzata)
Per $\delta > 0$, si ha

$$\log\big(\Gamma(z)\big) = \left(z - \frac{1}{2}\right)\log(z) - z + \frac{1}{2}\log(2\pi) + O(|z|^{-1}),$$

per $|z| \to +\infty$ nell'angolo $|\arg(z)| \leq \pi - \delta$.

In esso, $\delta > 0$ e $|\arg(z)| \leq \pi - \delta$ garantisce l'analiticità di tale relazione: infatti, se l'argomento diviene $-\pi$ (dunque per $z \in \mathbb{R}^-$), la precedente scrittura non è più valida (a causa della particolarità del logaritmo complesso (§3.2.10)).

8.4 La funzione $\prod$

In origine era stata proposta una notazione alternativa per quanto riguarda la funzione Gamma ([3]). Fu introdotta da Gauss e lo stesso Riemann la usò nella sua ricerca.
Essa fa riferimento alla funzione $\prod$ che, riferita alla Γ, si traduce nel modo seguente:

$$\prod(z) = \Gamma(z+1)$$

e dunque

$$\prod(n) = n!, \qquad n \in \mathbb{N}.$$

In termini della funzione $\prod$, la formula di riflessione si traduce con:

$$\prod(z)\prod(-z) = \frac{\pi z}{\sin(\pi z)}.$$

Inoltre una rappresentazione sotto forma di prodotto ricavabile dal limite di Gauss è

$$\prod(z) = \prod_{n=1}^{\infty} \frac{n^{1-z}(n+1)^z}{z+n} = \prod_{n=1}^{\infty} \frac{(1+1/n)^z}{1+z/n},$$

deducibile ricordando che $\prod(z) = \Gamma(z+1)$.

Questa forma, sebbene in gran parte non utilizzata, è considerata ([9], §1.3) più elegante e più naturale proprio a causa della "semplice" formula di riflessione e del fatto che $\prod(n) = n!$.

9. IL LOGARITMO INTEGRALE

In questa piccola sezione parleremo di una particolare funzione che ha un ruolo importante all'interno della teoria dei numeri. Essa, inoltre, ricorrerà anche in approssimazioni più accurate riguardo alla stima dei numeri primi minori di un intero dato e, quindi, della funzione $\pi(x)$.

9.1 Il logaritmo integrale

Il logaritmo integrale viene solitamente introdotto in almeno due modi. Entrambi fanno in qualche modo riferimento a integrali, il che giustifica il nome che la funzione riceve. Si definisce, anzitutto, il logaritmo integrale nel modo seguente:

$$li(x) = \int_0^x \frac{dt}{\log(t)}.$$

Tuttavia, la scrittura stessa crea subito qualche problema per $x \geq 1$ poiché l'integrando possiede una singolarità per $t = 1$ (si veda anche la Figura 9.1) in quanto il logaritmo al denominatore si annulla in quel punto. L'integrale, dunque, va interpretato nel modo seguente ([23], §1.5; [6])

$$li(x) = \begin{cases} \displaystyle\int_0^x \frac{dt}{\log(t)}, & 0 < x < 1 \\ \displaystyle\lim_{\varepsilon \to 0} \left(\int_0^{1-\varepsilon} \frac{dt}{\log(t)} + \int_{1+\varepsilon}^x \frac{dt}{\log(t)} \right), & x > 1 \end{cases}.$$

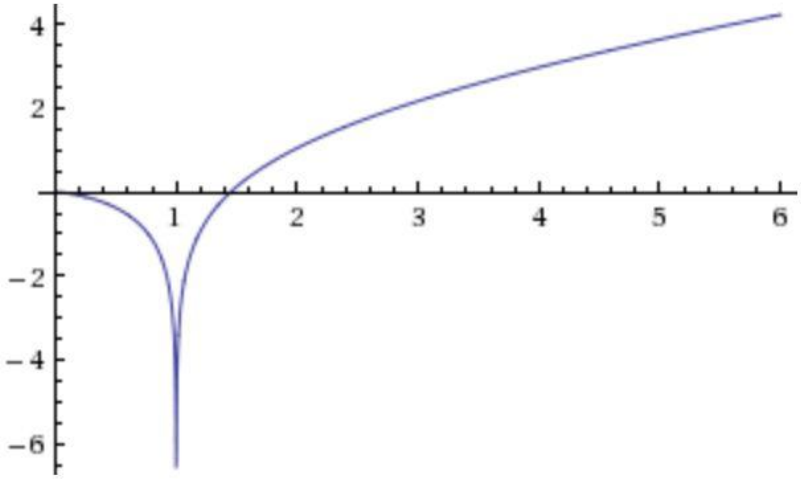

Figura 9.1. Grafico della funzione li per $0 < x < 6$ *(da wolframalpha.com* [33]*).*

Usualmente, allora, si fa riferimento a un'altra nozione di logaritmo integrale, proposta da Eulero e utilizzata in seguito anche da Gauss ([13], §1.1.6; [6]; [23], §1.5):

$$Li(x) = li(x) - li(2) = \int_2^x \frac{dt}{\log(t)}, \qquad x \geq 2.$$

Quest'ultima definizione consente, appunto, di evitare la singolarità nella quale si incorre quando $t = 1$ nell'integrando.

Tuttavia la notazione non è molto chiara poiché in vari testi risulta invertita: in ([23], §1.5), ([6]), ([22], §5.5) seguono la stessa logica utilizzata in questo paragrafo per introdurre il logaritmo integrale ($li(x)$ per quello generico, $Li(x)$ per quello Euleriano) mentre in altri testi, ad es. ([3], §4) e ([13], §1.1.6) accade esattamente l'opposto.

Per quanto ci riguarda, salvo controindicazioni, in futuro intenderemo per $Li(x)$ il logaritmo integrale Euleriano.

Prima di elencare alcune proprietà dei logaritmi integrali, occorre fare un'osservazione tanto semplice quanto importante. Dal teorema fondamentale del calcolo integrale è banale notare che i logaritmi integrali sono funzioni derivabili: $li(x)$ è derivabile $\forall x \neq 1$ mentre $Li(x)$ è derivabile in tutto il suo dominio (essendo definito per $x \geq 2$). Prendendo, ad esempio, la funzione $li(x)$, per il teorema fondamentale del calcolo integrale

$$\frac{d}{dx} li(x) = \frac{1}{\log(x)}$$

e un ragionamento simile vale per la funzione $Li(x)$.

Dunque, per il teorema di continuità risulta che entrambe le funzioni, nei punti in cui sono derivabili, sono anche continue.

Vediamo, ora, alcune proprietà dei logaritmi integrali:

(i) $Li(x) = \frac{x}{\log(x)} + \int_2^x \frac{dt}{\log(t)} - \frac{2}{\log(2)}$.

(ii) $li(2) = 1{,}045163780117492784$ ... (costante di Ramanujan-Soldner).

(iii) $li(\mu) = 0$, per $\mu = 1{,}4513692348$ ... (costante di Soldner).
Notiamo che questo valore μ esiste per il teorema di esistenza dei valori intermedi per una funzione ad una variabile. Infatti, siccome la funzione $li(x)$ è continua e si vede che $li(x) \to -\infty$ per $x \to 1^+$, si deduce per il teorema della permanenza del segno $li(x) < 0$ in un intorno destro di 1. Inoltre $li(2) > 0$ per la (ii). La conclusione è, dunque, l'esistenza di un valore μ tale che $li(\mu) = 0$.
L'unicità di questo valore è garantita dal fatto che la derivata della funzione $li(x)$ – definita in precedenza – è positiva per $x > 1$, quindi la funzione $li(x)$ è monotòna crescente per $x > 1$.
La difficoltà, nel corso dei secoli, fu proprio il calcolo del valore numerico di μ.

(iv) $\int_0^1 li(x)\, dx = \log(2)$.

Valgono, inoltre, altri risultati interessanti. Li elenchiamo e commentiamo senza entrare nei dettagli delle dimostrazioni.

Formula di Nielsen-Ramanujan

$$li(x) = \gamma + \log(log(x)) + \sum_{k=1}^{\infty} \frac{\log^k(x)}{k! \cdot k}.$$

Si intende qui che γ è la costante di Eulero-Mascheroni esaminata nella sezione dedicata (§7).

Una dimostrazione di questo risultato si ottiene semplicemente integrando per parti la funzione $li(x)$ così come è stata definita.

Vediamo, dunque, di operare l'integrazione per parti dal punto di vista puramente meccanico (quindi analizzando l'integrale indefinito) dimenticandoci, per ora, della singolarità e degli estremi di integrazione. In realtà ci sarebbero questioni tecniche da analizzare più profondamente (come ad esempio la singolarità per $t = 1$ nell'integrando) che però non approfondiremo nel dettaglio.
Il punto di partenza è il seguente:

$$\int \frac{dt}{\log(t)} = \int \frac{e^x}{x} dx.$$

Per compiere questo primo passo si opera il cambio di variabile $x = \log(t)$ da cui $e^x = t$ e $e^x dx = dt$. Tuttavia la funzione iniziale, così come quella ottenuta con il cambio di variabile, non è elementarmente integrabile, cioè non esiste una composizione di funzioni elementari che rappresenta il $li(x)$ evitando la sua definizione data sotto forma di integrale.
Ci si serve allora dello sviluppo in serie dell'esponenziale (§1.3.2)

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}, \qquad \Rightarrow \frac{e^x}{x} = \frac{\sum_{k=0}^{\infty} \frac{x^k}{k!}}{x} = \sum_{k=0}^{\infty} \frac{x^{k-1}}{k!}.$$

A questo punto basta integrare sfruttando il teorema di integrazione per serie (§2.2.2)

$$\int \frac{e^x}{x} dx = \int \sum_{k=0}^{\infty} \frac{x^{k-1}}{k!} dx = \sum_{k=0}^{\infty} \int \frac{x^{k-1}}{k!} dx = \int \frac{1}{x} dx + \sum_{k=1}^{\infty} \frac{x^k}{k! \cdot k} + c$$
$$= \log(x) + \sum_{k=1}^{\infty} \frac{x^k}{k! \cdot k} + c.$$

Occorre fare un paio di considerazioni prima di concludere.

- La costante c deriva dal calcolo dell'integrale indefinito.
- La sommatoria sotto il segno di integrale è stata divisa in due parti per la linearità dell'integrale stesso, isolando il caso $k = 0$ che si può facilmente integrare a parte (ottenendo $\log(x)$):

$$\sum_{k=0}^{\infty} \frac{x^{k-1}}{k!} = \frac{1}{x \cdot 0!} + \sum_{k=1}^{\infty} \frac{x^{k-1}}{k!} = \frac{1}{x} + \sum_{k=1}^{\infty} \frac{x^{k-1}}{k!}.$$

A questo punto, operiamo il cambio di variabile inverso, cioè torniamo alla variabile t (ricordando $x = \log(t)$, per ottenere la formula definitiva:

$$\int \frac{dx}{\log(x)} = \log(\log(t)) + \sum_{k=1}^{\infty} \frac{\log^k(t)}{k! \cdot k} + c.$$

Qui termina la parte "semplice" della dimostrazione, quella che abbiamo voluto mostrare in questo paragrafo come applicazione delle formule per il calcolo integrale. Tuttavia la dimostrazione vera richiede di considerare gli estremi di integrazione – quindi l'integrale definito – e, applicati alla formula appena trovata, dimostrare che $c = \gamma$, la costante di Eulero Mascheroni.

Formula di Ramanujan ([6]; [25])

$$li(x) = \gamma + \log(\log(x)) + \sqrt{x} \sum_{n=1}^{\infty} \frac{(-1)^{n+1} \log^n(x)}{n! \cdot 2^{n-1}} + \sum_{k=0}^{\lfloor (n-1)/2 \rfloor} \frac{1}{2k+1}.$$

Questa formula, al di là della sua apparente difficoltà, dà un'approssimazione migliore rispetto a quella del teorema precedente troncandola per un preciso valore di n (anziché ∞). In essa γ è la costante di Eulero-Mascheroni mentre $\lfloor (n-1)/2 \rfloor$ è la parte intera del numero $(n-1)/2$.

9.2 Il logaritmo integrale e i numeri primi

La funzione $Li(x)$ consente un'approssimazione più accurata della distribuzione dei primi rispetto alla precedente stima asintotica congetturata da Gauss (ed in seguito dimostrata indipendentemente da J. Hadamard e Ch. De La Vallée-Poussin (§6.1.11, §Appendice III)). I teoremi che andiamo a enunciare ci consentiranno un confronto tra i valori di $\pi(x)$, quelli di $x/\log(x)$ e $Li(x)$.
Tuttavia anche questa nuova stima è stata formulata dal matematico Gauss.

Teorema
Per $x \to \infty$,

$$li(x) = O\left(\frac{x}{\log(x)}\right).$$

Per essere più precisi ([6]),

$$li(x) \sim \frac{x}{\log(x)} \sum_{k=0}^{\infty} \frac{k!}{\log^k(x)}.$$

Risultati più accurati si ottengono con la funzione $Li(x)$.

Teorema ([23], §1.5)

$$\lim_{x\to\infty} \frac{\pi(x)}{Li(x)} = 1.$$

Questo risultato è da confrontare con la tabella riassuntiva di alcuni valori di $\pi(x), x/\log(x)$ e $Li(x)$ che esporremo alla fine di questa sezione..

Teorema ([22], §5.5)
Sia $\kappa(x) = (\log(x))^{3/5}(\log(\log(x)))^{-1/5}$, allora per qualche $c > 0$ si ha

$$\pi(x) = Li(x) + O\left(xe^{-c\kappa(x)}\right).$$

Congettura ([22], §5.5)
Per ogni $x > 2{,}01$, si ha

$$|\pi(x) - Li(x)| < x^{1/2}\log(x).$$

Quest'ultima affermazione è in realtà equivalente all'ipotesi di Riemann. Concludiamo con qualche esempio che illustra come il logaritmo integrale euleriano approssima nella pratica la funzione $\pi(x)$.

x	$\pi(x)$	$x/\log(x)$	$Li(x)$	$\pi(x) \Big/ \frac{x}{\log(x)}$	$\frac{\pi(x)}{Li(x)}$
10^2	25	21,7	30	1,15	0,83
10^3	168	144,9	178	1,16	0,94
10^4	1229	1086	1246	1,11	0,986
10^5	9592	8686	9630	1,10	0,996
10^6	78498	72464	78628	1,08	0,9983
10^7	664579	621118	664918	1,07	0,99949
10^8	5761455	5434780	5762209	1,06	0,999869
10^9	50847534	48309180	50849235	1,0525	0,999966

In questa tabella si può vedere come l'approssimazione di $\pi(x)$ mediante $Li(x)$ è decisamente più efficace (per valori maggiori di 1000) rispetto a quella vista in precedenza (§6.1.11)). Tuttavia il logaritmo integrale non è solo una mera approssimazione della funzione enumerativa dei primi: vedremo come esso ricoprirà un ruolo importante anche all'interno dell'ipotesi di Riemann.

10. TEORIA ANALITICA DEI NUMERI

In questa sezione tratteremo alcuni importanti concetti di una branca particolarmente interessante e complicata della Teoria dei Numeri qual è la Teoria Analitica dei Numeri. Nella Teoria Analitica dei Numeri – che potremo anche abbreviare con TADN – convergono idee e metodi dell'analisi matematica (reale e complessa) per trattare problemi riguardanti gli interi.
A questo punto non ci si dovrebbe più stupire nel vedere come ambiti piuttosto distanti – quali l'analisi e le questioni concernenti i numeri interi – si intreccino nel creare nuovi risultati che aprono mondi differenti impensabili in precedenza.

10.1 LE FUNZIONI ARITMETICHE

Il primo passo da compiere riguarda l'analisi di una classe particolare di funzioni, cioè le funzioni che sono definite sugli interi positivi ma ammettono valori reali o complessi. Le chiameremo aritmetiche. In questa sottosezione ne forniremo una descrizione concisa, insieme alle loro proprietà fondamentali. Forniremo anche qualche esempio pratico analizzando delle funzioni che meritano un rispetto particolare all'interno della Teoria Analitica dei Numeri.

10.1.1 Alcuni esempi famosi di funzioni aritmetiche: $\boldsymbol{\phi}, \boldsymbol{\mu}, \boldsymbol{\Lambda}$

Ribadiamo che una funzione a valori reali (o anche complessi) definita sugli interi positivi è detta funzione aritmetica ([3], §2.1). Proponiamo qualche esempio di funzione aritmetica, in parte già noto e comunque storicamente famoso e utile per sviluppi futuri.

La funzione $\boldsymbol{\phi}$ di Eulero
Cominciamo dalla funzione di Eulero ϕ, già trattata in precedenza (§6.2.6). Quindi ci limitiamo qui a rivederla in breve.
Per n intero positivo, avevamo definito $\phi(n)$ nel modo seguente

$$\phi(n) = \text{numero degli interi } a, 0 < a < n, \text{tali che } (a, n) = 1.$$

Tra le sue varie proprietà ricordiamo le più importanti:

- $\phi(p) = p - 1$, più in generale $\phi(p^k) = p^{k-1}(p-1)$ per p numero primo e k intero positivo;
- $\phi(n \cdot m) = \phi(n) \cdot \phi(m)$, per $(n, m) = 1$.

Con l'uso di queste due proprietà è infatti possibile costruire un algoritmo che determina il valore di $\phi(n)$ per ogni n intero $n > 1$, basato sulla decomposizione di n in fattori primi (con tutte le difficoltà che ne conseguono). Vale poi ovviamente $\phi(1) = 1$.

La funzione di Möbius ([3], §2.2)

Il modo in cui viene introdotta la funzione di Möbius è assai tecnico e poco intuitivo. Tuttavia la funzione ricorre in formule molto complesse ed è spesso utile a proposito della ζ di Riemann. La si indica con la lettera μ. La si definisce ponendo anzitutto

$$\mu(1) = 1.$$

Se poi $n > 1$, si decompone nel prodotto di potenze di fattori primi distinti come

$$n = \prod_{i=1}^{r} p_i^{a_i} = p_1^{a_1} \cdot p_2^{a_2} \cdot \ldots \cdot p_r^{a_r}, \qquad p_i \text{ primo}, \qquad a_i \geq 1,$$

si pone

$$\mu(n) = \begin{cases} (-1)^r, & \text{se } a_1 = a_2 = \cdots = a_r = 1, \\ 0, & \text{altrimenti}. \end{cases}$$

In altre parole

- se i fattori primi della decomposizione di n sono tutti distinti, $\mu(n)$ vale $+1$ o -1 a seconda che il numero di questi fattori sia pari o dispari,
- se invece c'è almeno un fattore primo che ricorre due volte, $\mu(n) = 0$.

Vediamo di fare un paio di esempi pratici.

Consideriamo, inizialmente, $n = 24$: $24 = 2^3 \cdot 3$ dunque $\mu(24) = 0$ poiché compare un fattore primo con un esponente ≥ 2 (in questo caso 2^3).

Se, invece, $n = 91$, $91 = 7 \cdot 13$ dunque $\mu(91) = (-1)^2 = 1$.

Inoltre, se n è un qualsiasi numero primo, allora $\mu(n) = -1$ in quanto la fattorizzazione di n è un banale $n = n$ (quindi n possiede un unico fattore con esponente 1).

Teorema

Per $n \geq 1$, risulta

$$\sum_{d|n} \mu(d) = I(n) := \left[\frac{1}{n}\right] = \begin{cases} 1, & \text{se } n = 1 \\ 0, & \text{se } n > 1 \end{cases}.$$

La funzione $I(n)$ è detta "funzione identità", ma non va confusa con la funzione identica, quella che lascia fisso ogni n (cioè $f(n) = n$).

Teorema ([3], §2.4)

Per $n \geq 1$, abbiamo

$$\phi(n) = \sum_{d|n} \mu(d) \frac{n}{d}.$$

Questo risultato ci mostra come a partire dalla funzione μ si possano anche dedurre i valori della funzione ϕ. Tuttavia neppure ricorrendo a μ e alla sua definizione il calcolo di ϕ diventa agevole perché continua a presupporre la conoscenza della fattorizzazione di n (o almeno del numero e della molteplicità dei suoi fattori).

La funzione di von Mangoldt ([3], §2.8)
Per ogni intero $n \geq 1$, definiamo

$$\Lambda(n) = \begin{cases} \log(p), & \text{se } n = p^m \text{ per qualche } p \text{ primo e } m \geq 0, \\ 0, & \text{altrimenti.} \end{cases}$$

Contrariamente alla funzione μ, il calcolo della Λ è più agevole poiché esistono algoritmi rapidi per riconoscere le potenze perfette dei numeri interi positivi, in particolare dei numeri primi (al contrario di quanto avviene per la fattorizzazione).
Vediamo una tabella riassuntiva che confronta alcuni valori delle funzioni di Möbius e di von Mangoldt.

n	1	2	3	4	5	6	7	8	9	10
$\mu(n)$	1	-1	-1	0	-1	1	-1	0	0	1
$\Lambda(n)$	0	$\log 2$	$\log 3$	$\log 2$	$\log 5$	0	$\log 7$	$\log 2$	$\log 3$	0

Teorema
Per $n \geq 1$, si ha

$$\log(n) = \sum_{d|n} \Lambda(d).$$

Teorema
Per $n \geq 1$, abbiamo

$$\Lambda(n) = \sum_{d|n} \mu(d) \log\left(\frac{n}{d}\right) = -\sum_{d|n} \mu(d) \log(d).$$

10.1.2 Prime proprietà delle funzioni aritmetiche

Date due funzioni aritmetiche f e g chiamiamo prodotto di convoluzione (o di Dirichlet) di f e g la funzione aritmetica $h = f * g$ definita nel modo seguente: per ogni intero positivo n

$$h(n) = (f * g)(n) = \sum_{d|n} f(d) g\left(\frac{n}{d}\right) = \sum_{d_1 d_2 = n} f(d_1) g(d_2).$$

L'ultima uguaglianza è facile da verificare ponendo $d = d_1$ e $n/d = d_2$ al variare di d, d_1, d_2.
Per esempio calcoliamo il prodotto di Dirichlet tra ϕ di Eulero e la μ di Möbius per $n = 12$. Ricordiamo che i divisori di 12 sono 1,2,3,4,6,12. Vale allora:

$$
\begin{aligned}
(\phi * \mu)(12) &= \sum_{d|n} \phi(d)\mu\left(\frac{n}{d}\right) \\
&= \phi(1)\mu\left(\frac{12}{1}\right) + \phi(2)\mu\left(\frac{12}{2}\right) + \phi(3)\mu\left(\frac{12}{3}\right) + \phi(4)\mu\left(\frac{12}{4}\right) + \phi(6)\mu\left(\frac{12}{6}\right) \\
&+ \phi(12)\mu\left(\frac{12}{12}\right) \\
&= \phi(1)\mu(12) + \phi(2)\mu(6) + \phi(3)\mu(4) + \phi(4)\mu(3) + \phi(6)\mu(2) \\
&+ \phi(12)\mu(1) = 1 \cdot 0 + 1 \cdot 1 + 2 \cdot 0 + 2 \cdot (-1) + 2 \cdot (-1) + 4 \cdot 1 \\
&= 1 - 2 - 2 + 4 = 1.
\end{aligned}
$$

Teorema ([3], §2.6)
Il prodotto di Dirichlet è commutativo e associativo, in altre parole scelte tre qualsiasi funzioni aritmetiche f, g, k abbiamo

- $f * g = g * f$ (proprietà commutativa);
- $(f * g) * k = f * (g * k)$ (proprietà associativa).

La prima proprietà è facile da verificare (si può far fede all'esempio visto prima del teorema) mentre la seconda è un po' più complicata.

Teorema ([3], §2.6)
Per ogni funzione aritmetica f abbiamo $f * I = I * f = f$.

Ricordiamo che I era la funzione identità vista nel paragrafo precedente.

10.1.3 Inverse e formula di inversione di Möbius

Ci si può allora chiedere se ogni funzione aritmetica ammetta una inversa, che sia anch'essa una funzione aritmetica rispetto al prodotto di Dirichlet e alla funzione identità I.

Teorema ([3], §2.7)
Sia una funzione aritmetica tale che $f(1) \neq 0$, allora esiste un'unica funzione aritmetica f^{-1}, detta anche inversa di Dirichlet di f, tale che

$$f * f^{-1} = f^{-1} * f = I.$$

Inoltre, se f, g sono due funzioni aritmetiche (tali che $f(1), g(1) \neq 0$) allora

$$(f * g)^{-1} = f^{-1} * g^{-1}.$$

Consideriamo ora la funzione unità u, ovvero la funzione aritmetica che assume sempre valore 1, $u(n) = 1, \ \forall n$. Il teorema precedente visto in (§10.1.1) ci dice che

$$\sum_{d|n} \mu(d) = I(n),$$

che si può esprimere affermando

$$\mu * u = I.$$

Questo implica che l'inversa di μ è la funzione unità u e viceversa:

$$\mu = u^{-1}, \qquad u = \mu^{-1}.$$

Teorema (Formula di Inversione di Möbius) ([3], §2.7; [28], §2.1)
Siano f e g due funzioni aritmetiche. Se

$$f(n) = \sum_{d|n} g(d),$$

allora

$$g(n) = \sum_{d|n} f(d)\mu\left(\frac{n}{d}\right).$$

Da notare che con "formula di inversione", non si intende una formula per trovare l'inversa di una funzione aritmetica ma una formula per "invertire la posizione" di due funzioni aritmetiche coinvolte nella relazione esposta nel teorema appena visto.
Gli ultimi due teoremi appena visti si possono generalizzare a risultati ancora più ampi: essi non saranno trattati all'interno di questa sezione poiché non inerenti agli obiettivi di questa tesi. La loro comprensione, inoltre, presuppone conoscenze più avanzate rispetto a quanto è stato trattato ora. Per chi è interessato, si rimanda alla lettura di ([3], §2.14) o anche ([28], §2.1).

10.1.4 Funzioni moltiplicative

In questo paragrafo descriveremo brevemente una classe molto interessante all'interno delle funzioni aritmetiche. Si tratta delle così dette funzioni moltiplicative. Ne sono esempi la ϕ, la μ e la I. Prima di verificarlo, però, diamo la definizione di funzione moltiplicativa.
Una funzione aritmetica è detta moltiplicativa se essa non è identicamente nulla e, per ogni scelta di interi positivi n e m primi tra loro, si ha

$$f(nm) = f(n)f(m), \qquad (n,m) = 1.$$

Se vale addirittura

$$f(nm) = f(n)f(m), \qquad \forall n, m,$$

allora la funzione è detta completamente moltiplicativa.
Possiamo ora confermare gli esempi già segnalati.

- La funzione ϕ di Eulero è moltiplicativa ma non completamente moltiplicativa.
  Infatti
  $$\phi(4) = \phi(2^2) = 2 \neq \phi(2) \cdot \phi(2) = 1 \cdot 1 = 1.$$
  Ricordiamo infatti che $\phi(p^k) = p^{k-1}(p-1)$ per p primo e $k \geq 1$.
- La funzione μ è moltiplicativa ma non completamente moltiplicativa.
  Infatti due interi positivi primi tra loro n e m non possono condividere fattori primi. Dunque se il prodotto nm ha un fattore primo multiplo, questo accade già per n o per m, così che $\mu(nm) = 0 = \mu(n) \cdot \mu(m)$; se invece il prodotto nm non ha fattori primi

multipli, è perché né m né n li hanno, e il numero dei fattori primi nel prodotto coincide con la somma di quelli di n e m rispettivamente: da questo si deduce facilmente che vale di nuovo $\mu(nm) = \mu(n)\mu(m)$. In questo modo si prova che la funzione μ è moltiplicativa.
Per convincersi che non è completamente moltiplicativa basta considerare, ad esempio, p^2 dove p un primo qualsiasi. Infatti

$$\mu(p^2) = 0 \neq \mu(p) \cdot \mu(p) = (-1) \cdot (-1) = 1.$$

- La funzione identità $I(n)$ è completamente moltiplicativa (anche perché $I(n) = 0$ per ogni $n \neq 1$).

Teorema ([3], §2.9)
Se f è una funzione moltiplicativa, allora $f(1) = 1$.

Da notare che la funzione di van Mangoldt non è moltiplicativa poiché $\Lambda(1) = 0$.

Teorema ([3], §2.10)
Se f e g sono funzioni moltiplicative, allora $f * g$ è anch'essa moltiplicativa. Inoltre se f e $f * g$ sono moltiplicative, lo è anche g.

Da questo teorema possiamo osservare – con un linguaggio non strettamente matematico – che il prodotto di Dirichlet conserva la "moltiplicatività". Tuttavia se f e g sono completamente moltiplicative, non è detto che $f * g$ sia completamente moltiplicativa (pur essendo moltiplicativa).

Teorema ([3], §2.11)
Sia f una funzione moltiplicativa. Allora f è completamente moltiplicativa se e solo se

$$f^{-1}(n) = \mu(n)f(n), \qquad \text{per ogni } n \geq 1.$$

Teorema ([3], §2.11)
Se f è moltiplicativa, allora

$$\sum_{d|n} \mu(d)f(d) = \prod_{p|n} \big(1 - f(p)\big).$$

Questo teorema, di per sé, non assume molta importanza all'interno di questa tesi, tuttavia in esso trova spazio un'idea ricorrente all'interno delle funzioni aritmetiche e delle serie di Dirichlet (che vedremo tra poco). L'idea è quella di trasformare una somma coinvolgente dei generici naturali in un prodotto nel quale compaiono solamente degli indici primi. La più importante formula di questo tipo sarà quella che riguarda il prodotto di Eulero.

10.1.5 Altre funzioni (moltiplicative)

Vediamo altri esempi di funzioni aritmetiche che hanno una certa importanza nella TADN ma meritano per noi interesse particolare in vista del loro coinvolgimento in varie rappresentazioni della funzione ζ di Riemann.

La funzione $\boldsymbol{\lambda(n)}$ di Liouville

La funzione λ di Liouville è un esempio di funzione completamente moltiplicativa.
Definiamo $\lambda(1) = 1$ mentre, per $n > 1$, scomposto nel prodotto di potenze di fattori primi distinti come

$$n = \prod_{i=1}^{r} p_i^{a_i}, \qquad p_i \text{ primo}, \qquad a_i \geq 1,$$

poniamo

$$\lambda(n) = (-1)^{a_1+a_2+\cdots+a_n}.$$

Teorema ([3], §2.12)

Per ogni $n \geq 1$, si ha

$$\sum_{d|n} \lambda(d) = \begin{cases} 1, & \text{se } n \text{ è un quadrato,} \\ 0, & \text{altrimenti.} \end{cases}$$

Inoltre $\lambda^{-1}(n) = |\mu(n)|$ per ogni $n \geq 1$.

Con $\lambda^{-1}(n)$ si intende l'inversa di λ rispetto al prodotto di convoluzione.

Le funzioni dei divisori $\boldsymbol{\sigma_\alpha(n)}$

Introduciamo un'altra classe di funzioni aritmetiche definite al variare di un parametro α; esse sono tutte moltiplicative. Definiamo, dunque, per α reale (o complesso) e per $n \geq 1$ intero

$$\sigma_\alpha(n) = \sum_{d|n} d^\alpha,$$

cioè la somma delle α −esime potenze dei divisori di n.
Possiamo notare che per $\alpha = 0, \sigma_0(n)$ è il numero dei divisori di n ($d^0 = 1$, per ogni d). Per $\alpha = 1$, invece, $\sigma_1(n)$ è la somma dei divisori di n, in questo caso la si indica anche con $S(n)$.

Teorema ([3], §2.13)

Se p è un numero primo e $k \geq 1$ è un intero, allora

$$\sigma_\alpha(p^k) = \begin{cases} \dfrac{p^{\alpha(k+1)} - 1}{p^\alpha - 1}, & \alpha \neq 0, \\ k + 1, & \alpha = 0. \end{cases}$$

Inoltre, per $n \geq 1$ si ha

$$\sigma_\alpha^{-1}(n) = \sum_{d|n} d^\alpha \mu(d) \mu\left(\frac{n}{d}\right),$$

nella quale σ_α^{-1} è l'inversa di σ_α rispetto al prodotto di Dirichlet.

10.1.6 Derivata di una funzione aritmetica e formula del prodotto di Eulero

Per ogni funzione aritmetica f, definiamo la sua derivata f' nel modo che segue

$$f'(n) = f(n)\log(n), \qquad n \geq 1.$$

E' banale notare che anch'essa è una funzione aritmetica.
Ad esempio, per quanto riguarda la funzione identità, $I'(n) = I(n)\log(n) = 0$ per ogni scelta di $n \geq 1$. Infatti l'unico valore per cui $I(n) \neq 0$ è per $n = 1$ ma in quel caso $\log(1) = 0$.

Teorema (proprietà della derivata) ([3], §2.18)
Siano f e g due funzioni aritmetiche. Risulta:

(i) $(f+g)' = f' + g'$;
(ii) $(f * g) = f' * g + f * g'$;
(iii) $(f^{-1})' = -f' * (f * f)^{-1}$, con $f(1) \neq 0$.

Teorema (identità di Selberg) ([3], §2.19)
Per $n \geq 1$, abbiamo

$$\Lambda(n)\log(n) + \sum_{d|n} \Lambda(d)\Lambda\left(\frac{n}{d}\right) = \sum_{d|n} \mu(d)\log^2\left(\frac{n}{d}\right).$$

Teorema (prodotto di Eulero) ([21], §2.4; [28], §4.3)
Sia f una funzione aritmetica moltiplicativa tale che $\sum_{n\geq 1}|f(n)|$ è convergente. Vale, allora, la formula di Eulero

$$\sum_{n=1}^{\infty} f(n) = \prod_{p \text{ primo}} \frac{1}{(1 - f(p))},$$

nel quale l'ultimo è il prodotto esteso a tutti i numeri primi p.

Questo teorema è molto importante e meriterebbe anche spazio una sua dimostrazione; tuttavia la ometteremo per la sua lunghezza anche se, a chi fosse interessato, si consiglia in particolare la lettura di ([21], §2.4).

Le seguenti osservazioni mettono in luce l'importanza di questo teorema.

Osservazione 1.
Consideriamo la seguente serie

$$\sum_{n=1}^{\infty} \frac{1}{n^s},$$

nella quale s è un numero complesso fissato. Possiamo notare che per $s \in \mathbb{R}$, essa non è altro che la serie armonica generalizzata già vista nella sezione di richiami di Analisi Matematica I (§1.2.4). Vedremo nel prossimo paragrafo, ma soprattutto nelle prossime sezioni, che questa non è una serie qualunque, ma la funzione ζ di Riemann definita al variare di s tale che $Re(s) > 1$ come

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \qquad s \in \mathbb{C}, \qquad Re(s) > 1.$$

Torneremo ampiamente sulla funzione $\zeta(s)$ nelle sezioni successive (essendo la funzione ζ di Riemann l'oggetto che è alla base di tutta la tesi), tuttavia, grazie alla formula del prodotto di Eulero, possiamo sin da ora affermare

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ primo}} \frac{1}{1 - \frac{1}{p^s}},$$

uguaglianza che tornerà utile nelle seguenti sezioni.

Osservazione 2 ([21], §2.4).
E' interessante notare che, grazie al teorema del prodotto di Eulero, si ha che la serie

$$\frac{1}{p_1} + \frac{1}{p_2} + \cdots + \frac{1}{p_n} + \cdots, \qquad p_i \text{ primo}, \qquad i \in \mathbb{N}$$

è divergente. Se, infatti, fosse convergente, allora lo sarebbe anche

$$\prod_{p \text{ primo}} \left(1 - \frac{1}{p}\right) = P,$$

nel quale abbiamo indicato con P il valore a cui converge il prodotto (infinito). Allora anche

$$\prod_{n \text{ primo}} \frac{1}{1 - \frac{1}{p}} = \frac{1}{P}$$

lo è (e converge a $1/P$). Quindi avremo che

$$\sum_{n=1}^{\infty} \frac{1}{n} = \prod_{p \text{ primo}} \frac{1}{1 - \frac{1}{p}} = \frac{1}{P}.$$

Invece sappiamo che il primo termine diverge essendo una serie armonica (§1.2.4).
Questa osservazione porta a conclusioni altrettanto importanti.

- I numeri primi sono infiniti poiché, se fossero finiti, diciamo p_i per $i \in \{1, \ldots, k\}$, la somma
$$\frac{1}{p_1} + \frac{1}{p_2} + \cdots + \frac{1}{p_n} + \cdots + \frac{1}{p_k},$$
sarebbe convergente perché finita. Questa è la dimostrazione di Eulero al teorema di Euclide sull'infinità dei numeri primi (§5.1.3).
- La stessa dimostrazione conferma che la serie armonica (semplice) è divergente. Attaccando il problema da un punto di vista alternativo, vedremo una dimostrazione del fatto che la serie armonica semplice, intesa come $\zeta(1)$, diverge in quanto $\zeta(s)$ opportunamente estesa ha un polo semplice per $s = 1$.

10.2 SERIE DI DIRICHLET

Passeremo, ora, ad analizzare le serie di Dirichlet.

Come vedremo a breve, anche la funzione ζ di Riemann è un esempio di una serie di Dirichlet.
Considereremo anche un risultato molto importante che è loro collegato e prende il nome di Formula della somma di Eulero. Di questa, vedremo brevemente alcune conseguenze legate alla funzione ζ di Riemann che saranno approfondite e richiamate più nel dettaglio nelle sezioni successive.

10.2.1 Serie di Dirichlet

Se consideriamo una qualsiasi successione $(a_n)_{n\in\mathbb{N}\setminus\{0\}}$ a valori reali (o complessi), definiamo la serie di Dirichlet associata nel modo seguente

$$f(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}.$$

In essa, la funzione f si dice funzione generatrice della successione a_n ([28], §2.4), s è la variabile (complessa) e n^s è il valore principale della potenza, cioè $n^s = e^{s\log(n)}$. Ricordiamo, infatti, che la potenza a valori complessi è una funzione a più valori mentre $\log(n)$ è il logaritmo (reale) del numero naturale n.
Ci si può chiedere come mai la TADN riserva tanto interesse a un "oggetto" che ha tutte le sembianze di una serie di funzioni di variabile complessa e che, almeno in apparenza, non ha molto a che vedere con altri argomenti di Teoria dei Numeri. La risposta è "storica": le serie di Dirichlet sono state introdotte in analisi per dimostrare un teorema famoso, dovuto appunto a Dirichlet sull'esistenza di infiniti numeri primi nella successione $an + b$ (§6.1.8), con a, b interi e $(a, b) = 1$.
La trattazione di questo tipo di serie non è la stessa dell'analisi complessa (o reale): sarà uno studio *misto* volto a mettere in luce proprietà particolari e forme differenti senza servirsi degli strumenti propri dell'analisi. Le serie di Dirichlet sono più utili nello studio dei numeri primi se introdotte nel loro appropriato contesto analitico ([28], §2.4).
Assegnata, dunque, la successione a_n, la somma della serie è una funzione della variabile complessa $s = \sigma + it$ (σ e t reali). E' di questa funzione vogliamo mettere in evidenza alcune proprietà fondamentali. Generalmente si utilizza la scrittura $s = \sigma + it$ (la stessa di Riemann, tra l'altro) invece della solita $z = x + iy$ per indicare la variabile complessa $s \in \mathbb{C}$.
Possiamo, ovviamente, cambiare punto di vista. Considerando la successione a_n come tipo particolare di funzione aritmetica, possiamo concludere che ad ogni funzione aritmetica si può associare una successione di Dirichlet.
Per esempio, alla funzione identità $I(n)$, si può associare la serie di Dirichlet

$$f(s) = \sum_{n=1}^{\infty} \frac{I(n)}{n^s} = 1,$$

infatti $I(n) \neq 0$ solo per $n = 1$.
Analogamente, alla funzione unità $u(n)$, si può associare la serie di Dirichlet

$$f(s) = \sum_{n=1}^{\infty} \frac{u(n)}{n^s} = \sum_{n=1}^{\infty} \frac{1}{n^s} = \zeta(s),$$

che è la zeta di Riemann (definita per $Re(s) > 1$) che analizzeremo nel dettaglio nelle sezioni seguenti e che abbiamo introdotto nel paragrafo precedente come osservazione al prodotto di Eulero.

Teorema (Jensen) ([21], §17.1)
Se la serie di Dirichlet $f(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}$ converge per s_0, converge anche per $Re(s) > Re(s_0)$.

Questo teorema è molto simile al lemma di Abel visto nella sezione di Analisi Complessa (§3.2.5), anche se l'oggetto è differente. In quel caso, infatti, si parlava di serie di potenze mentre quelle che trattiamo ora sono semplicemente esempi di serie di funzioni (ma non di potenze). Esso ci dice che se la serie di Dirichlet converge per s_0, converge anche in tutto il semipiano a destra di s_0 (cioè $Re(s) > Re(s_0)$). In generale, si può mostrare ([21], §17.1) che esiste un s_0 tale per cui la serie di Dirichlet converge per $Re(s) > Re(s_0)$ e non converge per $Re(s) < s_0$: in questo caso s_0 è detto ascissa di convergenza semplice.

Teorema ([21], §17.1)
Sia s_0 l'ascissa di convergenza semplice per la serie di Dirichlet $f(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}$. Allora

$$f(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s},$$

è olomorfa per $Re(s) > Re(s_0)$.

Questo risultato dimostra, in particolare, che la funzione

$$f(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \zeta(s)$$

definita per $Re(s) > 1$ e convergente per $Re(s) > 1$ è anche olomorfa per $Re(s) > 1$.

10.2.2 Formula di somma di Eulero

Teorema (formula di somma di Eulero) ([3], §3.3)
Sia f una funzione di classe C^1 in un intervallo reale $[a, b]$, in cui $0 < a < b$, allora

$$\sum_{a<n\leq b} f(n) = \int_a^b f(t)dt + \int_a^b (t - \lfloor t \rfloor) f'(t)dt + f(b)(\lfloor b \rfloor - b) - f(a)(\lfloor a \rfloor - a).$$

Analizzeremo meglio questa formula servendoci di un esempio. Ricordiamo che la scrittura $\lfloor x \rfloor$ sta a significare la parte intera di x (per x reale), cioè il più grande intero m tale che $m \leq x$; in altre parole $m \in \mathbb{Z}, m \leq x < m + 1$.

Consideriamo allora la funzione $f(x) = x^2$ per $x \in \left[\frac{1}{2}, \frac{5}{2}\right]$ che è di classe C^1 (in realtà C^∞) nell'intervallo $\left[\frac{1}{2}, \frac{5}{2}\right]$:

$$\sum_{\frac{1}{2}<n\leq\frac{5}{2}} n^2 = \int_{\frac{1}{2}}^{\frac{5}{2}} t^2 dt + \int_{\frac{1}{2}}^{\frac{5}{2}} (t - \lfloor t \rfloor)(2t) dt + \left(\frac{5}{2}\right)^2 \left(\left\lfloor \frac{5}{2} \right\rfloor - \frac{5}{2}\right) - \left(\frac{1}{2}\right)^2 \left(\left\lfloor \frac{1}{2} \right\rfloor - \frac{1}{2}\right).$$

Questa è la formula al completo, vediamo di analizzarla termine a termine. Anzitutto

$$\sum_{\frac{1}{2}<n\leq\frac{5}{2}} n^2 = 1^2 + 2^2 = 5.$$

In realtà, la prima sommatoria è quella che "in teoria" non conosciamo e quella per cui il teorema appena enunciato ci fornisce un modo alternativo di calcolo. Tuttavia, in questo esempio, abbiamo scelto una funzione volutamente semplice per far vedere, in pratica, come vale il teorema stesso. Da notare che la sommatoria è quella solita definita per indici interi positivi, infatti la scrittura

$$\sum_{a<n\leq b} f(n)$$

equivale a dire

$$\sum_{n=\lfloor a \rfloor+1}^{\lfloor b \rfloor} f(n).$$

Vediamo, ora, di analizzare termine a termine, il secondo membro della formula nel caso della funzione $f(x) = x^2$ presa come esempio.

Il primo è un semplice integrale:

$$\int_{\frac{1}{2}}^{\frac{5}{2}} t^2 dt = \left.\frac{t^3}{3}\right|_{\frac{1}{2}}^{\frac{5}{2}} = \frac{1}{3} \cdot \left(\frac{5}{2}\right)^3 - \frac{1}{3} \cdot \left(\frac{1}{2}\right)^3 = \frac{125}{24} - \frac{1}{24} = \frac{124}{24} = \frac{31}{6}.$$

Il secondo integrale è il più difficile. In realtà esistono vari metodi per calcolarlo, ma qui vedremo una tecnica piuttosto intuitiva senza entrare in dettagli tecnici. Iniziamo con l'analizzare la funzione $g(x) = x - \lfloor x \rfloor$.

In modo abbastanza intuitivo possiamo notare che

$$g(x) = x - \lfloor x \rfloor \in [0,1[, \qquad x \in \mathbb{R},$$

proprio per come è definita la parte intera di x. Al crescere di x, infatti, ogni volta che x è intero la sua parte intera cresce di un'unità, così che la differenza con il valore vero resta sempre inferiore all'unità. La $g(x)$ così definita è la parte decimale (o frazionaria) di x, cioè la differenza tra x e la sua parte intera. La si può indicare anche con la scrittura $\{x\}$.

Possiamo, dunque, scomporre l'integrale nel seguente modo

$$\int_{\frac{1}{2}}^{\frac{5}{2}}(t-\lfloor t\rfloor)(2t)dt=\int_{\frac{1}{2}}^{1}(t-\lfloor t\rfloor)(2t)dt+\int_{1}^{2}(t-\lfloor t\rfloor)(2t)dt+\int_{2}^{\frac{5}{2}}(t-\lfloor t\rfloor)(2t)dt$$

$$=\int_{\frac{1}{2}}^{1}2t^2dt+\int_{1}^{2}2t(t-1)dt+\int_{2}^{\frac{5}{2}}2t(t-2)dt$$

$$=\int_{\frac{1}{2}}^{1}2t^2dt+\int_{1}^{2}2t^2dt-\int_{1}^{2}2tdt+\int_{2}^{\frac{5}{2}}2t^2dt-\int_{2}^{\frac{5}{2}}4tdt$$

$$=\int_{\frac{1}{2}}^{\frac{5}{2}}2t^2dt-\int_{1}^{2}2tdt-\int_{2}^{\frac{5}{2}}4tdt=\left.\frac{2t^3}{3}\right|_{\frac{1}{2}}^{\frac{5}{2}}-t^2|_1^2-2t^2|_2^{\frac{5}{2}}$$

$$=\frac{2}{3}\cdot\left(\frac{5}{2}\right)^3-\frac{2}{3}\cdot\left(\frac{1}{2}\right)^3-4+1-2\cdot\left(\frac{5}{2}\right)^2+2\cdot 4=\frac{250}{24}-\frac{2}{24}-3-\frac{25}{2}+8$$

$$=\frac{248}{24}+5-\frac{25}{2}=\frac{31}{3}+5-\frac{25}{2}=\frac{62+30-75}{6}=\frac{17}{6}.$$

Abbiamo scomposto quest'integrale proprio perché, ottenendo estremi interi e intervalli unitari, abbiamo potuto sostituire $\lfloor t\rfloor$ con il suo valore assunto.
Passiamo agli ultimi due termini della formula.

$$\left(\frac{5}{2}\right)^2\left(\left\lfloor\frac{5}{2}\right\rfloor-\frac{5}{2}\right)-\left(\frac{1}{2}\right)^2\left(\left\lfloor\frac{1}{2}\right\rfloor-\frac{1}{2}\right)=\frac{25}{4}\cdot\left(-\frac{1}{2}\right)-\frac{1}{4}\cdot\left(-\frac{1}{2}\right)=-\frac{25}{8}+\frac{1}{8}=-\frac{24}{8}=-3.$$

A questo punto possiamo fare la somma di tutti i termini al secondo membro ottenendo, finalmente,

$$\int_{\frac{1}{2}}^{\frac{5}{2}}t^2dt+\int_{\frac{1}{2}}^{\frac{5}{2}}(t-\lfloor t\rfloor)(2t)dt+\left(\frac{5}{2}\right)^2\left(\left\lfloor\frac{5}{2}\right\rfloor-\frac{5}{2}\right)-\left(\frac{1}{2}\right)^2\left(\left\lfloor\frac{1}{2}\right\rfloor-\frac{1}{2}\right)=\frac{31}{6}+\frac{17}{6}-3=\frac{48}{6}-3$$
$$=8-3=5,$$

che è uguale al risultato ottenuto con la sommatoria al primo termine e, in ultima analisi, conferma il teorema.

10.2.3 Applicazioni della formula di somma di Eulero ([3], §3.4)

Applichiamo l'ultimo risultato mostrando alcune conseguenze piuttosto interessanti che esso ha su determinate serie (di Dirichlet); in particolare, siamo interessati alla ζ di Riemann.
Possiamo, dunque, partire con il considerare la seguente serie ($x\geq 1$)

$$\sum_{n\leq x}\frac{1}{n},$$

che non è altro che la somma parziale della serie armonica (semplice) ristretta a $0<n\leq x$.
Applichiamo, a questa, la formula di somma di Eulero considerando la somma degli $1/n$ come la restrizione della funzione $1/x$ ai naturali. E' un utilizzo corretto del teorema poiché la funzione $1/x$ è di classe C^1 per $x>0$. Si ha allora

$$\sum_{n\leq x}\frac{1}{n}=\int_1^x\frac{dt}{t}-\int_1^x\frac{t-\lfloor t\rfloor}{t^2}dt+1-\frac{x-\lfloor x\rfloor}{x}=\log(t)|_1^x-\int_1^x\frac{t-\lfloor t\rfloor}{t^2}dt+1-\frac{x-\lfloor x\rfloor}{x}$$
$$=\log(x)-\int_1^x\frac{t-\lfloor t\rfloor}{t^2}dt+1-\frac{x-\lfloor x\rfloor}{x}$$
$$=\log(x)-\int_1^{+\infty}\frac{t-\lfloor t\rfloor}{t^2}dt+\int_x^{+\infty}\frac{t-\lfloor t\rfloor}{t^2}dt+1-\frac{x-\lfloor x\rfloor}{x}.$$

In questa serie di uguaglianze abbiamo utilizzato in particolare la proprietà di linearità dell'integrale

$$\int_1^x\frac{t-\lfloor t\rfloor}{t^2}dt=\int_1^{+\infty}\frac{t-\lfloor t\rfloor}{t^2}dt-\int_x^{+\infty}\frac{t-\lfloor t\rfloor}{t^2}dt.$$

Facciamo, ora, tendere $x\to+\infty$ per valutare la serie armonica nella sua interezza. Ricordiamo che, per $x\to+\infty$,

$$\frac{x-\lfloor x\rfloor}{x}=O\left(\frac{1}{x}\right)\xrightarrow{x\to+\infty}0,$$

in quanto $x-\lfloor x\rfloor\in[0,1[$, come detto in precedenza.

$$\lim_{x\to+\infty}\sum_{n\leq x}\frac{1}{n}=\sum_{n=1}^{\infty}\frac{1}{n}=\boldsymbol{\lim_{x\to+\infty}\log(x)}-\int_1^{+\infty}\frac{t-\lfloor t\rfloor}{t^2}dt+\int_x^{+\infty}\frac{t-\lfloor t\rfloor}{t^2}dt+1.$$

Il termine in grassetto è quello che fornisce l'ennesima dimostrazione della divergenza della serie armonica semplice. Infatti $\log(x)$ tende a $+\infty$ per $x\to+\infty$ quindi questo termine è rimarcato proprio a testimoniare che è un'altra prova della divergenza armonica semplice.
Inoltre, sempre dal fatto che $x-\lfloor x\rfloor\in[0,1[$ e $x^2\geq 1$ per $x\geq 1$, si ha

$$0\leq\int_x^{+\infty}\frac{t-\lfloor t\rfloor}{t^2}\leq\int_x^{+\infty}\frac{1}{t^2}dt=\frac{1}{x}.$$

A questo punto, anche questo termine è anch'esso $O\left(\frac{1}{x}\right)$ e per $x\to+\infty$, è infinitesimo. Possiamo, allora, concludere

$$\sum_{n=1}^{\infty}\frac{1}{n}=\lim_{x\to\infty}\log(x)-\int_1^{+\infty}\frac{t-\lfloor t\rfloor}{t^2}dt+1.$$

Ricordando la definizione della costante di Eulero-Mascheroni (§7.1), questa diventa

$$\lim_{x\to\infty}\left(\sum_{n\leq x}\frac{1}{n}-\log(x)\right)=\gamma=-\int_1^{+\infty}\frac{t-\lfloor t\rfloor}{t^2}dt+1,$$

cioè

$$\gamma=-\int_1^{+\infty}\frac{t-\lfloor t\rfloor}{t^2}dt+1.$$

Vediamo, ora, di generalizzare e vedere cosa accade nella funzione ζ di Riemann che, come già detto, non è altro che l'estensione al semipiano complesso della serie armonica generalizzata. In analogia al caso precedente, partiamo con la somma parziale.

$$\sum_{n\leq x}\frac{1}{n^s},\qquad s\in\mathbb{C},\qquad Re(s)>1.$$

Applichiamo, dunque, la formula di somma di Eulero, considerando $f(x)=1/x^{-s}$, per $s\in\mathbb{C}$ tale che $Re(s)>1$. Otteniamo

$$\sum_{n \leq x} \frac{1}{n^s} = \int_1^x \frac{1}{t^s} dt - s \int_1^x \frac{t - \lfloor t \rfloor}{t^{s+1}} dt + 1 - \frac{x - \lfloor x \rfloor}{x^s}$$
$$= \int_1^x \frac{1}{t^s} dt - s \int_1^{+\infty} \frac{t - \lfloor t \rfloor}{t^{s+1}} dt + s \int_x^{+\infty} \frac{t - \lfloor t \rfloor}{t^{s+1}} dt + 1 - \frac{x - \lfloor x \rfloor}{x^s}$$
$$= \frac{x^{1-s}}{1-s} - \frac{1}{1-s} - s \int_1^{+\infty} \frac{t - \lfloor t \rfloor}{t^{s+1}} dt + 1 + O\left(\frac{1}{x^s}\right)$$
$$= \frac{x^{1-s} - 1}{1-s} - s \int_1^{+\infty} \frac{t - \lfloor t \rfloor}{t^{s+1}} dt + 1 + O\left(\frac{1}{x^s}\right)$$

Qui possiamo notare delle operazioni già viste in precedenza come la scrittura

$$\int_1^x \frac{t - \lfloor t \rfloor}{t^{s+1}} dt = \int_1^{+\infty} \frac{t - \lfloor t \rfloor}{t^{s+1}} dt - \int_x^{+\infty} \frac{t - \lfloor t \rfloor}{t^{s+1}} dt$$

e

$$0 \leq \int_x^{+\infty} \frac{t - \lfloor t \rfloor}{t^{s+1}} \leq \int_x^{+\infty} \frac{1}{t^{s+1}} dt = \frac{x^{-s}}{s} = \frac{1}{s x^s}$$

(quest'ultimo termine inglobato, come in precedenza, nell'$O\left(\frac{1}{x^s}\right)$).

Facendo tendere $x \to +\infty$, i termini $O\left(\frac{1}{x^s}\right)$ tendono a zero come $\frac{1}{x^s}$, dunque otteniamo

$$\lim_{x \to +\infty} \sum_{n \leq x} \frac{1}{n^s} = \sum_{n=1}^{+\infty} \frac{1}{n^s} = \zeta(s) = -\frac{1}{1-s} - s \int_1^{+\infty} \frac{t - \lfloor t \rfloor}{t^{s+1}} dt + 1.$$

Torneremo, nella sezione dedicata alla ζ di Riemann su questo risultato e su altri, tuttavia possiamo notare anche ora che in base a questa formula la funzione $\zeta(s)$ possiede un polo semplice per $s = 1$.

10.2.4 Le funzioni di Chebyshev

Per $x > 0$ reale definiamo la funzione ψ di Chebyshev mediante la seguente formula

$$\psi(x) = \sum_{n \leq x} \Lambda(n).$$

Sempre per $x > 0$, definiamo la funzione ϑ di Chebyshev nel modo che segue

$$\vartheta(x) = \sum_{p \leq x} \log(p), \qquad p \text{ primo}.$$

A proposito della funzione ψ, si ha che la $\Lambda(n)$ non si annulla solo per $n = p^k$ con p primo e $k \geq 1$ (nei quali casi vale $\log(p)$) (§10.1.1). Possiamo, allora, riscrivere la definizione di $\psi(x)$ in un altro modo

$$\psi(x) = \sum_{n \leq x} \Lambda(n) = \sum_{\substack{m=1 \\ p^m \leq x}}^{\infty} \sum_{p \text{ primo}} \Lambda(p^m) = \sum_{m=1}^{\infty} \sum_{p \leq x^{1/m}} \log(p) = \sum_{m=1}^{\lfloor \log_2(x) \rfloor} \vartheta\left(x^{1/m}\right),$$

espressione che mostra come sono legate tra loro le due funzioni appena descritte. Nell'ultima uguaglianza si è troncata la somma a $\log_2(x)$. La seconda sommatoria, infatti, esiste solo se

$$x^{\frac{1}{m}} \geq 2,$$

perché 2 è il più piccolo numero primo. Si ottiene, dunque

$$\log\left(x^{\frac{1}{m}}\right) = \frac{1}{m}\log(x) \geq \log(2)$$

e quindi

$$\frac{\log(x)}{\log(2)} = \log_2(x) \geq m,$$

visto che si può passare direttamente alla maggiorazione con la parte intera di $\log_2(x)$ proprio poiché l'indice m è intero (positivo).

Teorema ([3], §4.2)
Per $x > 0$ si ha

$$0 \leq \frac{\psi(x)}{x} - \frac{\vartheta(x)}{x} \leq \frac{\log^2(x)}{2\sqrt{x}\log(2)},$$

la quale, per $x \to \infty$, implica

$$\lim_{x\to\infty}\left(\frac{\psi(x)}{x} - \frac{\vartheta(x)}{x}\right) = 0.$$

Dimostrazione
Daremo una semplice dimostrazione di questo teorema poiché esso è un tassello molto importante nell'Appendice III, dedicato al teorema dei numeri primi.
Dalla

$$\psi(x) = \sum_{m=1}^{\lfloor \log_2(x) \rfloor} \vartheta\left(x^{1/m}\right),$$

isolando il termine per $m = 1$ per poi portarlo al primo membro si ottiene

$$0 \leq \psi(x) - \vartheta(x) = \sum_{m=2}^{\lfloor \log_2(x) \rfloor} \vartheta\left(x^{1/m}\right).$$

Ora, dalla definizione di $\vartheta(x)$ segue la seguente catena di disuguaglianze (con p primo)

$$\vartheta(x) = \sum_{p \leq x} \log(p) \leq \sum_{p \leq x} \log(x) \leq x\log(x).$$

Così, sostituendole alla relazione ottenuta in precedenza, si ha

$$0 \leq \psi(x) - \vartheta(x) = \sum_{m=2}^{\lfloor \log_2(x) \rfloor} \vartheta\left(x^{1/m}\right) \leq \sum_{m=2}^{\lfloor \log_2(x) \rfloor} x^{\frac{1}{m}} \log\left(x^{\frac{1}{m}}\right) \leq \log_2(x)\, x^{\frac{1}{2}} \log\left(x^{\frac{1}{2}}\right)$$
$$= \frac{\log(x)}{\log(2)}\sqrt{x}\frac{1}{2}\log(x) = \frac{\sqrt{x}\log^2(x)}{2\log(2)},$$

nella quale si è utilizzato il fatto che $x > 0$ nel dire $x^{\frac{1}{m}} \leq x^{\frac{1}{2}}$. Dividendo per x ambo i membri si ottiene la tesi del teorema.

Oltre all'importanza del risultato appena ottenuto, nell'appendice dedicato alla dimostrazione del teorema dei numeri primi ne vedremo altri riguardanti le due funzioni di Chebyshev.

11. LA FUNZIONE ζ DI RIEMANN

In questa sezione inizieremo ad analizzare il fulcro di questa tesi, nonché l'oggetto "matematico" alla base dell'ipotesi di Riemann. Si tratta di una funzione particolare, dalle molteplici e variegate proprietà.
La definiremo a partire dalla serie armonica descrivendo le prime importanti proprietà che la legano più o meno direttamente alla Teoria dei Numeri (analitica e non). Tutte le proprietà che vedremo in questa sezione saranno una diretta conseguenza di risultati esaminati nelle scorse sezioni, salvo adeguati "ritocchi tecnici". Lo scopo è anche quello di dare ordine a tutte le rappresentazioni più o meno utili e caratteristiche della ζ.
Questa sezione è la prima *vera* sezione di questa tesi: inizieremo, infatti, ad analizzare le profondità dell'ipotesi di Riemann, l'oggetto della tesi stessa. Non parleremo, nello specifico, dell'ipotesi, ma, come detto, inizieremo ad entrare nei dettagli della funzione ζ, l'oggetto dell'ipotesi di Riemann.
In molte dimostrazioni ci serviremo delle proprietà della serie geometrica, estesa al piano complesso a partire da quella già vista nella sezione di richiami di Analisi I (§1.2.4)

$$\frac{1}{1-z} = \sum_{n=1}^{\infty} z^n, \qquad z \in \mathbb{C} \text{ tale che } |z| < 1.$$

11.1 Introduzione: dalla serie armonica generalizzata alla ζ

Consideriamo la seguente serie

$$\sum_{n=1}^{\infty} \frac{1}{n^x}, \qquad x \in \mathbb{R};$$

essa è la serie armonica generalizzata già introdotta nella sezione di richiami di Analisi Matematica I (§1.2.4). Avevamo anche visto che essa converge per $x > 1$ mentre diverge per $x \leq 1$.
Tuttavia la questione è più complicata: dal punto di vista di una serie di funzioni, si possono studiare – nell'intervallo di convergenza – le proprietà della funzione alla quale converge.
Per $x > 1$, sappiamo che la serie armonica generalizzata converge e, chiameremo $f(x)$ la funzione alla quale converge. Per x nell'intervallo di convergenza della serie armonica generalizzata, sommiamo a x un fissato valore $\varepsilon > 0$, così per $x \geq 1 + \varepsilon$, si ha $1/n^x \leq 1/n^{1+\varepsilon}$ per ogni n intero positivo, dunque

$$f(x) = \sum_{n=1}^{\infty} \frac{1}{n^x} \leq \sum_{n=1}^{\infty} \frac{1}{n^{1+\varepsilon}}, \qquad \text{per ogni } x \geq 1 + \varepsilon.$$

La relazione appena vista ci dice che la serie armonica generalizzata è dominata termine a termine da una serie numerica, quindi converge totalmente. Possiamo concludere che la

funzione alla quale converge è continua nell'intervallo $[1+\varepsilon, +\infty)$ proprio perché la convergenza totale implica quella uniforme (§2.2.1-2.2.2).
Tuttavia, con tecniche più raffinate – che qui non esponiamo in quanto non inerenti ai fini della tesi – si può anche concludere che $f(x)$ è di classe C^∞ nel suddetto intervallo e, anzi, che lo è in $(1, +\infty)$.
A questo punto possiamo estendere la funzione appena definita al campo complesso in maniera naturale. La funzione

$$f(s) = \zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \qquad Re(s) > 1$$

è l'unica estensione al semipiano complesso $Re(s) > 1$ della $f(x)$ definita in precedenza (§3.2.6). La funzione appena introdotta è la funzione ζ di Riemann che, quindi, non è altro che l'estensione al semipiano complesso $Re(s) > 1$ della funzione definita come somma della serie armonica generalizzata nei punti in cui quest'ultima converge.
Possiamo, quindi, definire la funzione ζ di Riemann nel modo seguente

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \qquad Re(s) > 1,$$

che non è altro che la serie di Dirichlet generata dalla funzione unità (§10.2.1). Come serie di Dirichlet, il punto $s = 1$ è l'ascissa di convergenza semplice per la serie e, per il teorema di Jensen (§10.2.1) si ha che $\zeta(s)$ è una funzione olomorfa per $Re(s) > 1$.

11.2 Alcune rappresentazioni della ζ

Grazie all'opera del matematico Eulero, si hanno le due rappresentazioni molto interessanti che abbiamo già accennato nella sezione di Teoria Analitica dei Numeri e che qui riprenderemo.
La prima è quella dovuta al teorema del prodotto di Eulero (§10.1.6)

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ primo}} \frac{1}{1 - \frac{1}{p^s}}$$

e ci fa notare come la ζ si colleghi – in un certo qual modo – ai numeri primi.
Tuttavia questa non è l'unica rappresentazione dovuta ad Eulero, la più famosa è infatti quella già trattata nella sezione di Teoria Analitica dei Numeri – più precisamente (§10.2.2) – per il calcolo completo.
Riprendiamo, però, la formula finale

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = -\frac{1}{1-s} - s\int_1^{\infty} \frac{t - \lfloor t \rfloor}{t^{s+1}} dt + 1.$$

Se poniamo $\{t\} = t - \lfloor t \rfloor$ e inglobiamo il segno meno al primo termine otteniamo

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \frac{1}{s-1} - s\int_1^{\infty} \frac{\{t\}}{t^{s+1}} dt + 1.$$

Il merito di Eulero è stato anche quello di aver calcolato ([1]) la somma

$$\zeta(2) = \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} \cong 1{,}6644934$$

ed inoltre di avere fornito (oltre ad altri valori concreti) una rappresentazione – che vedremo in seguito – per i valori di $\zeta(n)$ per n intero positivo pari

Tuttavia si può andare oltre, trovando una rappresentazione analoga a quella appena vista senza servirsi della formula di somma di Eulero. A tal proposito, ricordiamo che

$$\int_n^{\infty} \frac{s}{t^{s+1}} dt = -\frac{1}{t^s}\bigg|_n^{\infty} = \frac{1}{n^s},$$

e

$$\int_n^{n+1} \frac{s}{t^{s+1}} dt = -\frac{1}{t^s}\bigg|_n^{n+1} = -\frac{1}{(n+1)^s} + \frac{1}{n^s};$$

inoltre si vede facilmente (basta isolare il primo termine e riordinare il resto della serie) che

$$\sum_{n=1}^{\infty} n\left(\frac{1}{n^s} - \frac{1}{(n+1)^s}\right) = 1 + \sum_{n=1}^{\infty}\left(-n\left(\frac{1}{(n+1)^s} + (n+1)\frac{1}{(n+1)^s}\right)\right) = 1 + \sum_{n=1}^{\infty} \frac{1}{(n+1)^s}$$
$$= \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

Sostituendo nella definizione della $\zeta(s)$,

$$\begin{aligned}
\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} &= \sum_{n=1}^{\infty} n\left(\frac{1}{n^s} - \frac{1}{(n+1)^s}\right) = s\sum_{n=1}^{\infty} n\int_n^{n+1} \frac{dt}{t^{s+1}} = s\sum_{n=1}^{\infty}\int_n^{n+1} \frac{\lfloor t \rfloor}{t^{s+1}} dt \\
&= s\int_1^{\infty} \frac{\lfloor t \rfloor}{t^{s+1}} dt = s\int_1^{\infty} \frac{t - \{t\}}{t^{s+1}} dt = s\int_1^{\infty} \frac{t}{t^{s+1}} dt - s\int_1^{\infty} \frac{\{t\}}{t^{s+1}} dt \\
&= s\int_1^{\infty} \frac{1}{t^s} dt - s\int_1^{\infty} \frac{\{t\}}{t^{s+1}} dt = \frac{s}{(-s+1)t^{s-1}}\bigg|_1^{\infty} - s\int_1^{\infty} \frac{\{t\}}{t^{s+1}} dt \\
&= -\frac{s}{-s+1} - s\int_1^{\infty} \frac{\{t\}}{t^{s+1}} dt = \frac{s}{s-1} - s\int_1^{\infty} \frac{\{t\}}{t^{s+1}} dt \\
&= \frac{s-1+1}{s-1} - s\int_1^{\infty} \frac{\{t\}}{t^{s+1}} dt = 1 + \frac{1}{s-1} - s\int_1^{\infty} \frac{\{t\}}{t^{s+1}} dt.
\end{aligned}$$

Questa formula è analoga a quella trovata in precedenza grazie alla formula della somma di Eulero ma è ottenuta con un procedimento differente basato solo sull'analisi dei termini che compaiono nella sommatoria.

La rappresentazione della $\zeta(s)$, dunque, è la seguente

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \frac{1}{s-1} - s\int_1^{+\infty} \frac{\{t\}}{t^{s+1}} dt + 1,$$

ma il termine al secondo membro è definito per $Re(s) > 0$ poiché l'integrale è convergente (tenuto sempre conto che $s \neq 1$ per evitare che la prima frazione perda significato). Per l'unicità del prolungamento analitico (§3.2.6), possiamo concludere che questa formula è già un'estensione analitica della $\zeta(s)$ al semipiano $Re(s) > 0$, proprio perché per $Re(s) > 1$ si ottiene la funzione così come è stata definita in partenza.

Possiamo, poi, notare che in corrispondenza del punto $s = 1$ la funzione possiede un polo semplice (§3.4.3). Inoltre, per $\mathbb{C}\backslash\{1\}$, la funzione è olomorfa in quanto somma di termini olomorfi:

- $\frac{1}{s-1}$ è una funzione olomorfa (sempre per $s \neq 1$);
- l'integrale è olomorfo poiché integrale di una funzione olomorfa (§3.3.4).

Dunque la funzione è complessivamente meromorfa (§3.4.3) in quanto olomorfa tranne una singolarità di tipo polo.

11.3 La rappresentazione integrale

A fianco degli usuali modi di rappresentare la ζ con il prodotto di Eulero o con la formula di somma di Eulero, sussiste un altro tipo di rappresentazione della ζ mediante la funzione Γ e un integrale.

Teorema (rappresentazione integrale)
Per $Re(s) > 1$, la funzione zeta di Riemann può essere definita dall'integrale

$$\zeta(s) = \frac{1}{\Gamma(s)} \int_0^\infty \frac{t^{s-1}}{e^t - 1} dt.$$

Dimostrazione
Verificheremo che la formula è corretta partendo dalla tesi per risalire alla formulazione della $\zeta(s)$. Notiamo anzitutto

$$\frac{t^{s-1}}{e^t - 1} = \frac{t^{s-1}}{e^t}\left(\frac{1}{1 - 1/e^t}\right) = \frac{t^{s-1}}{e^t} \sum_{k=0}^{\infty} \frac{1}{e^{kt}} = t^{s-1} \sum_{k=0}^{\infty} \frac{1}{e^{kt} \cdot e^t} = t^{s-1} \sum_{k=0}^{\infty} \frac{1}{e^{(k+1)t}} = \sum_{k=1}^{\infty} \frac{t^{s-1}}{e^{kt}}.$$

Si è utilizzata, qui, la proprietà della serie geometrica

$$\frac{1}{1 - 1/e^t} = \sum_{k=0}^{\infty} \frac{1}{e^{kt}},$$

in quanto $1/e^t < 1$. Inoltre, nella penultima sommatoria si sono cambiati gli indici, identificando il nuovo indice sempre con k e sfruttando la proprietà di traslazione degli indici per le sommatorie (§1.2.2).
A questo punto, nell'integrale, otteniamo

$$\int_0^\infty \frac{t^{s-1}}{e^t - 1} dt = \int_0^\infty \sum_{k=1}^{\infty} \frac{t^{s-1}}{e^{kt}} dt = \sum_{k=1}^{\infty} \int_0^\infty \frac{t^{s-1}}{e^{kt}} dt.$$

Nell'ultimo integrale operiamo il cambio di variabile $y = kt$, così che $dy = kdt$. Otteniamo:

$$\frac{1}{\Gamma(s)}\int_0^\infty \frac{t^{s-1}}{e^t-1}dt = \frac{1}{\Gamma(s)}\sum_{k=1}^\infty \int_0^\infty \frac{t^{s-1}}{e^{kt}}dt = \frac{1}{\Gamma(s)}\sum_{k=1}^\infty \int_0^\infty e^{-kt}t^{s-1}dt$$
$$= \frac{1}{\Gamma(s)}\sum_{k=1}^\infty \int_0^\infty \frac{e^{-y}}{k}\left(\frac{y}{k}\right)^{s-1}dy = \frac{1}{\Gamma(s)}\sum_{k=1}^\infty \int_0^\infty \frac{1}{k^s}e^{-y}y^{s-1}dy$$
$$= \frac{1}{\Gamma(s)}\sum_{k=1}^\infty \frac{1}{k^s}\int_0^\infty e^{-y}y^{s-1}dy = \frac{1}{\Gamma(s)}\sum_{k=1}^\infty \frac{1}{k^s}\Gamma(s) = \frac{\Gamma(s)}{\Gamma(s)}\sum_{k=1}^\infty \frac{1}{k^s} = \sum_{k=1}^\infty \frac{1}{k^s}$$
$$= \zeta(s).$$

Ricordiamo, infatti, (§8,2) che, indici a parte,

$$\int_0^\infty e^{-y}y^{s-1}dy = \Gamma(s),$$

per la definizione stessa di funzione Gamma.

11.4 Un collegamento tra la ζ e i primi ([26], §1.1)

In questo paragrafo vedremo un ulteriore collegamento tra la funzione ζ e i numeri primi. In realtà, l'ipotesi di Riemann – se dimostrata (o meno) – avrà conseguente molto più profonde con i numeri primi e la Teoria dei Numeri, rispetto a quello che introdurremo qui.
Introduciamo, anzitutto, il seguente integrale

$$\int_n^{n+1} \frac{-s}{t(t^s-1)}dt = \int_n^{n+1} \frac{-s}{t^{s+1}(1-1/t^s)}dt.$$

Notiamo che l'ultimo integrale è della forma

$$\int_n^{n+1} \frac{1}{f(t)}f'(t)\,dt = \log\big(f(t)\big)\Big|_n^{n+1},$$

in cui $f(t) = 1 - 1/t^s$. Dunque

$$\int_n^{n+1} \frac{-s}{t^{s+1}\left(1-\frac{1}{t^s}\right)}dt = \log\left(1-\frac{1}{t^s}\right)\Big|_n^{n+1} = \log\left(1-\frac{1}{(n+1)^s}\right) - \log\left(1-\frac{1}{n^s}\right).$$

Ricordando la formula del prodotto di Eulero

$$\zeta(s) = \prod_{p \text{ primo}} \frac{1}{1-1/p^s},$$

possiamo, allora, trasformare la funzione $\zeta(s)$ nel modo seguente servendoci delle proprietà di sommatorie e prodotti infiniti viste in (§1.2.2)

$$\begin{aligned}
\log(\zeta(s)) &= \log\left(\prod_{p\ primo} \frac{1}{1-1/p^s}\right) = \sum_{p \text{ primo}} \log\left(\frac{1}{1-1/p^s}\right) \\
&= \sum_{p \text{ primo}} \left(\log(1) - \log\left(1-\frac{1}{p^s}\right)\right) = -\sum_{p \text{ primo}} \log\left(1-\frac{1}{p^s}\right) \\
&= -\sum_{n=2}^{\infty} \big(\pi(n)-\pi(n-1)\big)\log\left(1-\frac{1}{n^s}\right) \\
&= -\sum_{n=2}^{\infty} \pi(n)\log\left(1-\frac{1}{n^s}\right) + \sum_{n=2}^{\infty} \pi(n-1)\log\left(1-\frac{1}{n^s}\right) \\
&= -\sum_{n=2}^{\infty} \pi(n)\log\left(1-\frac{1}{n^s}\right) + \sum_{n=1}^{\infty} \pi(n)\log\left(1-\frac{1}{(n+1)^s}\right) \\
&= -\sum_{n=2}^{\infty} \pi(n)\log\left(1-\frac{1}{n^s}\right) + \sum_{n=2}^{\infty} \pi(n)\log\left(1-\frac{1}{(n+1)^s}\right) \\
&= -\sum_{n=2}^{\infty} \pi(n)\left(\log\left(1-\frac{1}{n^s}\right) - \log\left(1-\frac{1}{(n+1)^s}\right)\right) \\
&= \sum_{n=2}^{\infty} \pi(n)\int_n^{n+1} \frac{s}{t(t^s-1)}\,dt = s\int_2^{\infty} \frac{\pi(t)}{t(t^s-1)}\,dt.
\end{aligned}$$

In questa serie di uguaglianze, il passaggio da p primo a $n \geq 2$ è dovuto all'utilizzo della funzione $\pi(n)$ poiché $\pi(n)-\pi(n-1)=1$ solo se n è primo, altrimenti vale zero e tutto il termine della sommatoria è nullo, inoltre, $\pi(n)=0$ per $n<2$.

11.5 Collegamenti tra la ζ e alcune funzioni aritmetiche

In questo paragrafo vedremo alcuni modi alternativi di rappresentare la funzione ζ mediante altre funzioni aritmetiche o serie di Dirichlet.

Teorema
Per $Re(s)>1$, si ha

$$\frac{1}{\zeta(s)} = \prod_{p \text{ primo}} \left(1-\frac{1}{p^s}\right) = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}.$$

Di questo teorema non daremo la dimostrazione, poiché si richiedono dei risultati più avanzati che non sono stati trattati in questa tesi. Tuttavia la prima uguaglianza richiede proprietà già viste sui prodotti infiniti.

Oltre a questo, che è un risultato abbastanza classico, ve ne sono altri non meno importanti.

Teorema
Per $s \in \mathbb{C}$ tale che $Re(s) > 1$, risulta

$$\frac{\zeta'(s)}{\zeta(s)} = -\sum_{n=2}^{\infty} \frac{\Lambda(n)}{n^s}.$$

Dimostrazione
Si dimostra facendo qualche calcolo

$$\frac{\zeta'(s)}{\zeta(s)} = \frac{1}{\zeta(s)} \frac{d}{ds}\left(\prod_{p \text{ primo}} \frac{1}{1-p^{-s}}\right) = \frac{1}{\zeta(s)} \sum_{p \text{ primo}} \frac{d}{dz}\left(\frac{1}{1-p^{-s}}\right) \prod_{q \neq p \text{ primo}} \frac{1}{1-q^{-s}}$$

$$= \sum_{p \text{ primo}} \left(\frac{-p^{-s}\log(p)}{(1-p^{-s})^2}\right)(1-p^{-s}) = \sum_{p \text{ primo}} -\frac{\log(p)}{p^s}\left(1-\frac{1}{p^s}\right)^{-1}$$

$$= -\sum_{p \text{ primo}} \log(p) \sum_{m=1}^{\infty} \frac{1}{p^{ms}} = -\sum_{n=2}^{\infty} \frac{\Lambda(n)}{n^s}.$$

In essa si è utilizzata una generalizzazione della derivata del prodotto: ricordiamo, da alcune basilari proprietà dell'analisi matematica che

$$(f \cdot g)' = f'g + fg',$$

per f e g funzioni derivabili. Prendendo la scrittura appena vista come base, si può estendere questa proprietà (per induzione sull'indice n dello stesso prodotto) a prodotti infiniti convergenti ottenendo

$$\left(\prod_{n=1}^{\infty} f_n\right) = f_1' f_2 f_3 \ldots + f_1 f_2' f_3 \ldots + \cdots = \sum_{n=1}^{\infty}\left[f_n' \prod_{k=1, k\neq n}^{\infty} f_k\right].$$

Inoltre si è sostituito $\left(1-\frac{1}{p^s}\right)^{-1}$ con il suo sviluppo di Taylor, rifacendosi alla serie geometrica richiamata ad inizio sezione in quanto, per $s = \sigma + it \in \mathbb{C}$

$$\left|-\frac{1}{p^s}\right| = \frac{1}{|p^s|} = \frac{1}{|p^{\sigma+it}|} = \frac{1}{|p^\sigma||p^{it}|} = \frac{1}{p^\sigma|e^{it\log(p)}|} = \frac{1}{p^\sigma|e^{it}|^{\log(p)}} = \frac{1}{p^\sigma \cdot 1^{\log(p)}} = \frac{1}{p^\sigma} < 1.$$

In essa, per p primo, $p^\sigma > p > 1$ poiché si sta trattando la definizione di $\zeta(s)$ tramite il prodotto di Eulero ($Re(s) = \sigma > 1$). Infine, l'ultima uguaglianza è ottenuta sostituendo la prima sommatoria con la definizione della $\Lambda(n)$, che non si annulla solo se n è una potenza di un numero primo.

Teorema
Per $Re(s) > 1$ si ha

$$\zeta^2(s) = \sum_{n=1}^{\infty} \frac{\sigma_0(n)}{n^s}.$$

Ricordiamo (§10.1.5) che con $\sigma_\alpha(n)$ avevamo definito una classe di funzioni aritmetiche

$$\sigma_\alpha(n) = \sum_{d|n} d^\alpha,$$

al variare del parametro α reale (o complesso) con $n \geq 1$ intero.

Una semplice dimostrazione di questo fatto, la si può trovare in ([26], §1.2). Tuttavia essa si basa su un'importante proprietà del prodotto tra serie che non è stata esaminata in questa tesi. Vediamo, inoltre, un'altra collezione di risultati simile alla precedente.

Teorema ([26], §1.2)
Sia $\zeta(s)$ definita per $Re(s) > 1$, allora

(i) $\frac{\zeta(s)}{\zeta(2s)} = \sum_{n=1}^{\infty} \frac{|\mu(n)|}{n^s}$;

(ii) $\frac{\zeta^2(s)}{\zeta(2s)} = \sum_{n=1}^{\infty} \frac{2^{\nu(n)}}{n^s}$;

(iii) $\frac{\zeta^3(s)}{\zeta(2s)} = \sum_{n=1}^{\infty} \frac{\sigma_0(n^2)}{n^s}$;

(iv) $\frac{\zeta^4(s)}{\zeta(2s)} = \sum_{n=1}^{\infty} \frac{(\sigma_0(n))^2}{n^s}$.

Nella (ii), $\nu(n)$ è il numero dei fattori primi (diversi) di n.

Dimostrazione
Per la (ii), (iii), (iv) ci si serve delle proprietà della serie geometrica già utilizzate nella precedente dimostrazione:

$$\frac{1}{1-z} = \sum_{n=1}^{\infty} z^n, \qquad z \in \mathbb{C} \text{ tale che } |z| < 1.$$

Iniziamo, però, con il provare la (i), tenendo conto delle proprietà dei prodotti infiniti (§1.2.2).

$$\frac{\zeta(s)}{\zeta(2s)} = \left(\prod_{p \text{ primo}} \left(1 - \frac{1}{p^s}\right)^{-1}\right) \Bigg/ \left(\prod_{p \text{ primo}} \left(1 - \frac{1}{p^{2s}}\right)^{-1}\right) = \prod_{p \text{ primo}} \frac{1 - 1/p^{2s}}{1 - 1/p^s}$$
$$= \prod_{p \text{ primo}} \frac{(1 - 1/p^s)(1 + 1/p^s)}{1 - 1/p^s} = \prod_{p \text{ primo}} \left(1 + \frac{1}{p^s}\right),$$

da qui non è difficile concludere la tesi poiché il risultato è uguale a quello trovato per il teorema a inizio paragrafo.

Passiamo alla (ii), con analoghi passaggi e proprietà.

$$\frac{\zeta^2(s)}{\zeta(2s)} = \left(\prod_{p \text{ primo}} \left(1 - \frac{1}{p^s}\right)^{-2}\right) \Bigg/ \left(\prod_{p \text{ primo}} \left(1 - \frac{1}{p^{2s}}\right)^{-1}\right) = \prod_{p \text{ primo}} \frac{1 - 1/p^{2s}}{(1 - 1/p^s)^2}$$
$$= \prod_{p \text{ primo}} \frac{(1 - 1/p^s)(1 + 1/p^s)}{(1 - 1/p^s)^2} = \prod_{p \text{ primo}} \frac{1 + 1/p^s}{1 - 1/p^s}$$
$$= \prod_{p \text{ primo}} \frac{1 + 1/p^s - 2 + 2}{1 - 1/p^s} = \prod_{p \text{ primo}} \frac{-1 + 1/p^s + 2}{1 - 1/p^s}$$
$$= \prod_{p \text{ primo}} \left(-1 + \frac{2}{1 - 1/p^s}\right) = \prod_{p \text{ primo}} (-1 + 2 + 2p^{-s} + 2p^{-2s} + 2p^{-3s} + \ldots)$$
$$= \prod_{p \text{ primo}} (1 + 2p^{-s} + 2p^{-2s} + 2p^{-3s} + \ldots) = \prod_{p \text{ primo}} \left(\frac{1}{1 - 2/p^s}\right)$$
$$= \prod_{p \text{ primo}} \left(\frac{1}{1 - 2^1/p^s}\right) = \prod_{p \text{ primo}} \left(\frac{1}{1 - 2^{\nu(p)}/p^s}\right) = \sum_{n=1}^{\infty} \frac{2^{\nu(n)}}{n^s}.$$

Quest'ultima si ottiene applicando la formula del prodotto di Eulero (§10.1.6) in senso inverso, passando cioè dal prodotto infinito alla sommatoria. Inoltre, essendo $\nu(p)$ il numero dei fattori primi distinti di p, nella produttoria, essendo p primo, $\nu(p) = 1$.
Per la (iii) e la (iv) il procedimento è analogo e si arriva, in entrambi i casi a formulazioni del tipo

$$\frac{\zeta^3(s)}{\zeta(2s)} = \prod_{p \text{ primo}} \frac{1 + 1/p^s}{(1 - 1/p^s)^2}$$

per la (iii) e

$$\frac{\zeta^4(s)}{\zeta(2s)} = \prod_{p \text{ primo}} \frac{1 + 1/p^s}{(1 - 1/p^s)^3}$$

per la (iv) dalle quali, applicando più volte le regole della serie geometrica (con opportuni coefficienti come nella (ii)) e il prodotto di Eulero, si giunge alla relazione desiderata.

Il seguente teorema, invece, ci dà delle relazioni tra la funzione $\zeta(s)$ e altre funzioni aritmetiche.

Teorema ([26], §1.2)
Sia $\zeta(s)$ definita per $Re(s) > 1$, allora

(i) $\frac{\zeta(2s)}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\lambda(n)}{n^s}$,

(ii) $\frac{\zeta(s-1)}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\phi(n)}{n^s}$.

Dimostrazione
La dimostrazione segue il filo logico di quella precedente, con una serie di calcoli espliciti seguiti dall'applicazione delle formule per la serie geometrica e, infine, del prodotto di Eulero per giungere alla tesi desiderata.
Dimostriamo, ad esempio, la (i).

$$\frac{\zeta(2s)}{\zeta(s)} = \left(\prod_{p \text{ primo}} \left(1 - \frac{1}{p^{2s}}\right)^{-1}\right) \Bigg/ \left(\prod_{p \text{ primo}} \left(1 - \frac{1}{p^s}\right)^{-1}\right) = \prod_{p \text{ primo}} \frac{1 - 1/p^s}{1 - 1/p^{2s}}$$

$$= \prod_{p \text{ primo}} \frac{1 - 1/p^s}{(1 - 1/p^s)(1 + 1/p^s)} = \prod_{p \text{ primo}} \frac{1}{1 + p^{-s}} = \sum_{n=1}^{\infty} \left(-\frac{1}{n^s}\right) = \sum_{n=1}^{\infty} \frac{\lambda(n)}{n^s},$$

quest'ultima ricordando la definizione della funzione $\lambda(n)$ (§10.1.5).

Teorema ([26], §1.2)
Sia $\sigma_\alpha(n)$ la somma delle α −esime potenze dei divisori di n così come è stata definita in (§10.1.5). Allora, per $Re(s) > 1$

$$\zeta(s)\zeta(s - \alpha) = \sum_{n=1}^{\infty} \frac{\sigma_\alpha(n)}{n^s}.$$

12. PROLUNGAMENTI ANALITICI DELLA FUNZIONE ζ

Nella sezione precedente, avevamo definito la funzione

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s},$$

per $Re(s) > 1$. Essa è la funzione ζ di Riemann che, di per sé, è già un'estensione al semipiano complesso $Re(s) > 1$ della serie armonica generalizzata.
Tuttavia, così com'è definita, questa funzione diverge per $Re(s) \leq 1$. Lo scopo di questa sezione sarà di estenderla a tutto il piano complesso – prima a $Re(s) > 0$ ($s \neq 1$), poi a tutto $\mathbb{C}\backslash\{1\}$ – ad una funzione analitica $f(s)$ tale che:

- $f(s) = \zeta(s)$, per $Re(s) > 1$ (altrimenti non sarebbe un'estensione!);
- $f(s)$ non sia divergente come $\zeta(s)$ per $Re(s) \leq 1$ ($s \neq 1$).

Una volta trovata una funzione che soddisfa tali proprietà, per l'unicità del prolungamento analitico (§3.2.6) potremo concludere che questa è l'<u>unica</u> estensione analitica della $\zeta(s)$ a tutto il piano complesso (sempre con $s \neq 1$).
La sezione, dunque, è divisa in due parti.
Nella prima ci preoccuperemo di estendere la $\zeta(s)$ al semipiano $Re(s) > 0$ ($s \neq 1$). Nella seconda la estenderemo a tutto il piano complesso $\mathbb{C}\backslash\{1\}$ mentre nell'ultima dimostreremo – nei due modi utilizzati da Riemann – l'esistenza di un'equazione funzionale per la ζ.

12.1 PROLUNGAMENTO DELLA $\zeta(s)$ AL SEMIPIANO $Re(s) > 0$ ($s \neq 1$)

Inizieremo con l'estendere la funzione ζ di Riemann al semipiano $Re(s) > 0$ ($s \neq 1$).

12.1.1 Un primo passo

Nella precedente sezione, avevamo ripreso un risultato ottenuto applicando la formula della somma di Eulero (§10.2.2) alla funzione ζ. Eravamo giunti alla seguente uguaglianza

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = -\frac{1}{1-s} - s\int_1^{\infty} \frac{t - \lfloor t \rfloor}{t^{s+1}} dt + 1,$$

ovvero

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \frac{1}{s-1} - s\int_1^{\infty} \frac{\{t\}}{t^{s+1}}dt + 1.$$

Come avevamo già detto in precedenza questa era già un'estensione analitica della $\zeta(s)$ al semipiano complesso $Re(s) > 0$. Ovviamente $s \neq 1$ in quanto, come possiamo vedere dalla formula stessa, il valore $s = 1$ annullerebbe il denominatore al primo addendo e quindi la funzione non sarebbe complessivamente definita.

Per quanto riguarda il prolungamento analitico così ottenuto

- $\frac{1}{s-1}$ è una funzione olomorfa in tutto $Re(s) > 0$, tranne che in $s = 1$;
- $s\int_1^{\infty} \frac{t-\lfloor t \rfloor}{t^{s+1}}dt$ è convergente in quanto $Re(s) > 0$ ed è $O(t^{-s})$.

La funzione così ottenuta, dunque, per $Re(s) > 0$ ($s \neq 1$), è olomorfa (o analitica) e tale che, ristretta a $Re(s) > 1$ è proprio la $\zeta(s)$, dunque è il suo unico prolungamento analitico.

12.1.2 Un altro semplice prolungamento

Vediamo un altro modo per trovare un'estensione analitica della ζ alla regione $Re(s) > 0$.

Definiamo, innanzitutto, la funzione $\eta(s)$ – detta funzione η di Dirichlet – nel modo seguente:

$$\eta(s) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n^s}, \qquad Re(s) > 0.$$

Si può dimostrare che $\eta(s)$, definita come serie di Dirichlet, ha ascissa di convergenza semplice $s_0 = 0$. Analogamente a quanto accade per la funzione ζ di Riemann, anche la funzione η di Dirichlet possiede una sua rappresentazione integrale ([2]):

$$\eta(s) = \frac{1}{\Gamma(s)}\int_0^{\infty} \frac{t^{s-1}}{e^t + 1}dt,$$

che si può dimostrare in modo analogo a quanto fatto per la ζ, sviluppando l'integrale per poi tornare alla definizione originaria (§11.3) e che mostra molto più chiaramente la sua analiticità per $Re(s) > 0$.

Anche per la $\eta(s)$, un calcolo esplicito dei valori non è semplice. Tuttavia, per uno di essi, c'è una dimostrazione molto breve che si basa sullo sviluppo in serie di Taylor della funzione $\log(1 + x)$. Ricordiamo, infatti, che (§1.3.2)

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots + (-1)^{n+1}\frac{x^n}{n} + \cdots = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}x^n$$
$$= \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}x^n,$$

nella quale, nella sommatoria $(-1)^{n+1} = (-1)^{n-1+2} = (-1)^{n-1} \cdot (-1)^2 = (-1)^{n-1}$.

Ora, nel caso in cui $x = 1$, otteniamo

$$\log(1+1) = \log(2) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} = \eta(1).$$

Tuttavia la η non è una serie di Dirichlet qualsiasi. Essa fornisce un altro modo per prolungare la ζ alla regione $Re(s) > 0$, come dimostreremo con un procedimento semplice anche se non completamente rigoroso in termini matematici.
Per introdurlo ricordiamo ancora la definizione (classica) della funzione $\zeta(s)$

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \ldots, \qquad Re(s) > 1.$$

Moltiplichiamo ambo i membri per $1/2^s$

$$\frac{1}{2^s}\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{(2n)^s} = \frac{1}{2^s} + \frac{1}{4^s} + \frac{1}{6^s} + \frac{1}{8^s} + \ldots,$$

da cui, raddoppiando ambo i membri, si ricava

$$2\zeta(s) = \frac{1}{2^{s-1}}\zeta(s) = \sum_{n=1}^{\infty} \frac{2}{(2n)^s} = \frac{2}{2^s} + \frac{2}{4^s} + \frac{2}{6^s} + \frac{2}{8^s} + \ldots.$$

Infine aggiungiamo ad entrambi i membri la funzione $\eta(s)$ per poi concludere

$$\begin{aligned}\eta(s) + \frac{1}{2^{s-1}}\zeta(s) &= 1 - \frac{1}{2^s} + \frac{1}{3^s} - \frac{1}{4^s} + \frac{1}{5^s} - \frac{1}{6^s} + \frac{1}{7^s} - \frac{1}{8^s} + \cdots + \frac{2}{2^s} + \frac{2}{4^s} + \frac{2}{6^s} + \frac{2}{8^s} + \cdots \\ &= 1 + \left(-\frac{1}{2^s} + \frac{2}{2^s}\right) + \frac{1}{3^s} + \left(-\frac{1}{4^s} + \frac{2}{4^s}\right) + \frac{1}{5^s} + \left(-\frac{1}{6^s} + \frac{2}{6^s}\right) + \frac{1}{7^s} \\ &+ \left(-\frac{1}{8^s} + \frac{2}{8^s}\right) + \cdots = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \frac{1}{5^s} + \frac{1}{6^s} + \frac{1}{7^s} + \frac{1}{8^s} + \cdots = \sum_{n=1}^{\infty} \frac{1}{n^s} \\ &= \zeta(s),\end{aligned}$$

da cui concludiamo

$$\eta(s) = \zeta(s) - \frac{1}{2^{s-1}}\zeta(s) = \zeta(s)(1 - 2^{1-s}),$$

cioè

$$\zeta(s) = \frac{\eta(s)}{1 - 2^{1-s}}.$$

Quindi, poiché la funzione $\eta(s)$ è definita per $Re(s) > 0$, possiamo dedurre che la precedente espressione permette un prolungamento analitico della ζ al semipiano $Re(s) > 0$.
Inoltre l'unica singolarità è sempre $s = 1$ che, stavolta, annulla il denominatore.

12.2 ESTENSIONE A TUTTO $\mathbb{C}\backslash\{1\}$

Passiamo, ora, all'estensione della ζ a tutto il piano complesso (tranne $s = 1$). Troveremo varie forme per prolungare analiticamente la ζ a tutto $\mathbb{C}\backslash\{1\}$.
Nella sezione di Analisi Complessa (§3.2.6), avevamo visto che si tratta, dunque, di trovare una $f(s)$ analitica tale che

- $f(s) = \zeta(s)$, per $s \in \mathbb{C}$ tale che $Re(s) > 1$ (o anche $Re(s) > 0, s \neq 1$ considerando lo sviluppo che si ottiene tramite la formula di somma di Eulero,
- $f(s)$ è definita in tutto $\mathbb{C}\backslash\{1\}$.

A quel punto con abuso (giustificato) di notazione, potremo porre $f(s) \equiv \zeta(s)$ e con il termine "zeta di Riemann" intenderemo direttamente la funzione ampliata a tutto il piano complesso (tranne $s = 1$).

12.2.1 Un difficile integrale: l'estensione di Riemann ([26], §2.4)

In questo paragrafo analizzeremo la prima estensione della ζ a tutto $\mathbb{C}\backslash\{1\}$, operata da Riemann stesso (§Appendice I) tramite il calcolo di un integrale piuttosto complicato lungo un dominio di integrazione parecchio inusuale. Serviranno tutti i concetti dell'integrazione complessa su curve (§3.3).
Inizieremo, dunque, calcolando il seguente integrale

$$\int_{+\infty}^{+\infty} \frac{(-x)^s}{x(e^x - 1)} dx.$$

Se necessario possiamo porre (§3-2-11) $(-x)^s = e^{s \log(-x)}$.
A prima vista può sembrare davvero un integrale fuori da ogni qualsiasi logica: che vuol dire integrare da $+\infty$ a $+\infty$? Anche considerandolo come integrale improprio, non si riescono ad avere risposte migliori a tal senso.
Tuttavia stiamo operando in campo complesso e l'integrale si può ottenere come un integrale curvilineo

$$\int_{+\infty}^{+\infty} \frac{(-x)^s}{x(e^x - 1)} dx = \int_{\gamma} \frac{(-x)^s}{x(e^x - 1)} dx,$$

nel quale γ è la curva che parte da $+\infty$ lungo l'asse reale positivo, arriva ad incontrare la circonferenza $|x| = \delta$ (dove δ sarà introdotto nel giro di poche righe), la percorre in senso orario girando attorno all'origine, per poi tornare a $+\infty$ sempre lungo l'asse reale positivo come in Figura 12.1. Il raggio δ della circonferenza contenente lo zero deve rispettare $0 < \delta < 2\pi$ in modo che il cammino non incontri le discontinuità dell'integrando ma contenga l'origine (come dice Riemann stesso (§Appendice I)).

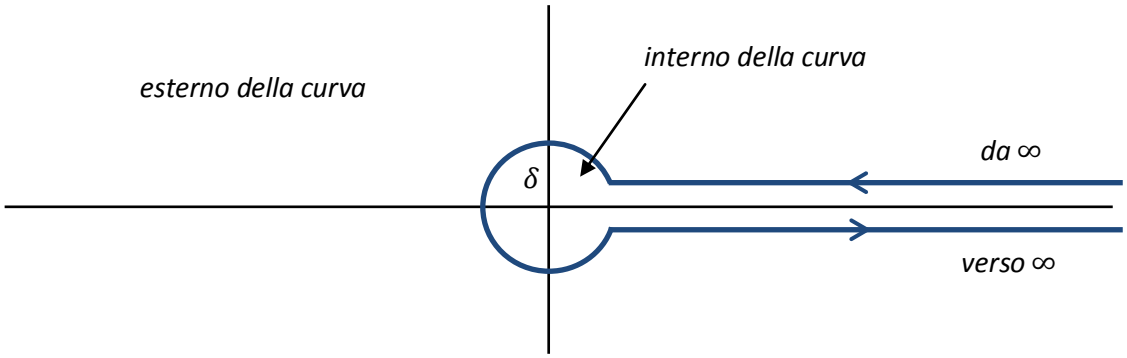

Figura 12.1. Contorno di integrazione.

Sebbene la Figura 12.1 sia esplicativa e largamente usata nella maggior parte dei testi che trattano più o meno approfonditamente dell'ipotesi di Riemann (come, ad esempio, ([21], §17.7)), tuttavia non è completamente corretta ma serve solamente per focalizzare il

problema. Il contorno di integrazione, infatti, *giace* sull'asse reale fino ad incontrare la circonferenza $|x| = \delta$ per poi percorrerla e ritornare all'asse reale fino a ∞ ([26], §2.4) (Figura 12.2).
Nella Figura 12.1 si deve allora pensare che la distanza della curva parallela all'asse reale sia trascurabile rispetto al raggio della circonferenza.

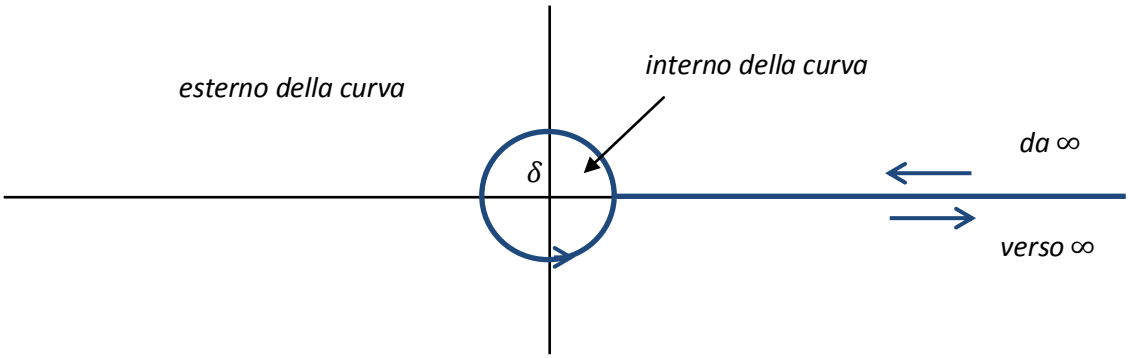

Figura 12.2. Contorno di integrazione.

In Figura 12.2 è mostrata la situazione con il cammino che giace sull'asse reale; le frecce indicano il verso di percorrenza (prima da $+\infty$ a incontrare la circonferenza, poi dalla circonferenza verso $+\infty$).
Si può notare che la curva è chiusa poiché composta da due cammini sovrapposti (senza alcun "interno") e una circonferenza intorno all'origine.
L'integrale, dunque, è "effettivamente" calcolato da $+\infty$ a $+\infty$ sebbene lungo un cammino insolito che include l'origine.
Procediamo con il calcolo esplicito: nel farlo, teniamo presente che il cammino di integrazione è una curva regolare a tratti poiché unione di 3 curve regolari, due rette e una circonferenza. Si ha

$$\int_{+\infty}^{+\infty} \frac{(-x)^s}{x(e^x-1)}dx = \int_{+\infty}^{\delta} \frac{(-x)^s}{(e^x-1)x}dx + \int_{|x|=\delta} \frac{(-x)^s}{(e^x-1)x}dx + \int_{\delta}^{+\infty} \frac{(-x)^s}{(e^x-1)x}dx.$$

Per $\delta \to 0$, il secondo integrale si annulla: non lo mostreremo poiché occorrono concetti avanzati di analisi complessa che non abbiamo trattato in questa tesi.
Passiamo, dunque, al resto del calcolo.

$$\begin{aligned}
\int_{+\infty}^{+\infty} \frac{(-x)^s}{x(e^x-1)}dx &= \lim_{\delta\to 0}\left(\int_{+\infty}^{\delta} \frac{(-x)^s}{(e^x-1)x}dx + \int_{\delta}^{+\infty} \frac{(-x)^s}{(e^x-1)x}dx\right) \\
&= \lim_{\delta\to 0}\left(\int_{+\infty}^{\delta} \frac{e^{s(\log(x)-i\pi)}}{(e^x-1)x}dx + \int_{\delta}^{+\infty} \frac{e^{(s\log(x)+i\pi)}}{(e^x-1)x}dx\right) \\
&= \lim_{\delta\to 0}\left(e^{-i\pi s}\int_{+\infty}^{\delta} \frac{e^{s\log(x)}}{(e^x-1)x}dx + e^{i\pi s}\int_{\delta}^{+\infty} \frac{e^{s\log(x)}}{(e^x-1)x}dx\right) \\
&= e^{-i\pi s}\int_{+\infty}^{0} \frac{x^s}{(e^x-1)x}dx + e^{i\pi s}\int_{0}^{\infty} \frac{x^s}{(e^x-1)x}dx \\
&= -e^{-i\pi s}\int_{0}^{+\infty} \frac{x^s}{(e^x-1)x}dx + e^{i\pi s}\int_{0}^{\infty} \frac{x^s}{(e^x-1)x}dx \\
&= \left(e^{i\pi s} - e^{-i\pi s}\right)\int_{0}^{+\infty} \frac{x^{s-1}}{e^x-1}dx.
\end{aligned}$$

Senza entrare in particolari dettagli tecnici, nel calcolo di questo integrale si è utilizzata la seguente proprietà: $(-x)^s = (-1)^s x^s$, dunque, passando all'esponenziale

$$(-x)^s = (-1)^s x^s = e^{s\log(-1)} e^{s\log(x)} = e^{\pm i\pi s} e^{s\log(x)} = e^{s(\pm i\pi + \log(x))}.$$

Il segno dell'esponente dipende dal ramo del logaritmo che si prende in considerazione (ricordando che il logaritmo complesso è una funzione a più valori, dunque ha più rami regolari (§3.2.10)): sopra all'asse reale, cioè da $+\infty$ a δ si considera il ramo corrispondente proprio al logaritmo reale mentre al di sotto dell'asse reale, da δ a $+\infty$, si prende quello reale meno 2π.

A questo punto ricordiamo la rappresentazione integrale della ζ di Riemann (§11.3)

$$\zeta(s) = \frac{1}{\Gamma(s)} \int_0^{\infty} \frac{t^{s-1}}{e^t - 1} dt.$$

L'ultimo integrale, a parte la variabile d'integrazione, è lo stesso presente nella forma integrale della ζ. Possiamo, dunque, sostituire

$$\int_{+\infty}^{+\infty} \frac{(-x)^s}{x(e^x - 1)} dx = \left(e^{i\pi s} - e^{-i\pi s}\right) \int_0^{+\infty} \frac{x^{s-1}}{e^x - 1} dx = \left(e^{i\pi s} - e^{-i\pi s}\right)\Gamma(s)\zeta(s)$$
$$= 2i\sin(\pi s)\,\Gamma(s)\zeta(s),$$

in essa, tramite la formula di Eulero (§3.2.7), si è posto

$$e^{i\pi s} - e^{-i\pi s} = \cos(\pi s) + i\sin(\pi s) - \cos(-\pi s) - i\sin(-\pi s)$$
$$= \cos(\pi s) + i\sin(\pi s) - \cos(\pi s) + i\sin(\pi s) = 2i\sin(\pi s).$$

Abbiamo, dunque, ottenuto la seguente uguaglianza

$$\int_{+\infty}^{+\infty} \frac{(-x)^s}{x(e^x - 1)} dx = 2i\sin(\pi s)\,\Gamma(s)\zeta(s).$$

Possiamo, però, andare oltre moltiplicando ambo i membri per

$$\frac{\Gamma(1-s)}{2\pi i}$$

ottenendo

$$\frac{\Gamma(1-s)}{2\pi i} \int_{+\infty}^{+\infty} \frac{(-x)^s}{x(e^x - 1)} dx = \frac{\Gamma(s)\Gamma(1-s)}{\pi} \sin(\pi s)\,\zeta(s) = \zeta(s),$$

dunque

$$\zeta(s) = \frac{\Gamma(1-s)}{2\pi i} \int_{+\infty}^{+\infty} \frac{(-x)^s}{x(e^x - 1)} dx.$$

A questo punto, possiamo fare le dovute osservazioni.

Si è utilizzato nel penultimo passaggio il principio di riflessione di Eulero per la funzione Γ (§8.3):

$$\Gamma(s)\Gamma(1-s) = \frac{\pi}{\sin(\pi s)}.$$

Ora, in questa rappresentazione della $\zeta(s)$, l'integrale converge sempre poiché l'esponenziale al denominatore domina una qualsiasi potenza della x al numeratore, inoltre la funzione, nei punti in cui è definita, è ovviamente analitica perché composizione di funzioni olomorfe. Possiamo, dunque, concludere che, quella appena trovata, è l'estensione della $\zeta(s)$ all'intero piano complesso $\mathbb{C}\backslash\{1\}$.

Per $s = 1$, la funzione Γ ha un polo semplice e quindi anche la funzione ζ ha un polo semplice, come si era già visto nella precedente rappresentazione dovuta alla formula della somma di Eulero. Tuttavia, la funzione $\Gamma(s)$ ha dei poli semplici per $s = 0, -1, -2, \ldots$,

dunque $\Gamma(1-s)$ li ha per $s = 1,2,\ldots$. Tuttavia, mentre per $s = 1$, il polo della Γ resta, per $s = 2,3,\ldots$, la funzione $\zeta(s)$ non ha poli, dunque si conclude che l'integrale ha degli zeri che cancellano i poli della Γ. Ulteriori chiarimenti di questo fatto, saranno evidenti parlando dell'equazione funzionale.
Se, ora, in questa rappresentazione

$$\zeta(s) = \frac{\Gamma(1-s)}{2\pi i}\int_{+\infty}^{+\infty} \frac{(-x)^s}{x(e^x-1)}dx,$$

usiamo la notazione che Riemann ha utilizzato per la funzione Γ, cioè

$$\Gamma(s+1) = \prod(s),$$

otteniamo

$$\zeta(s) = \frac{\prod(-s)}{2\pi i}\int_{+\infty}^{+\infty} \frac{(-x)^s}{x(e^x-1)}dx,$$

che è la stessa formula trovata da Riemann nel suo articolo di ricerca (si veda l'Appendice I per maggiori dettagli) utilizzando il procedimento illustrato in questo paragrafo.

12.2.2 Valori di $\zeta(s)$ per s intero negativo

La formula appena trovata, ci consente di fare ulteriori analisi. Tuttavia, prima di procedere oltre, conviene introdurre i numeri di Bernoulli.
Analizziamo la seguente funzione, strettamente legata all'integrando dell'equazione trovata a fine paragrafo precedente:

$$f(x) = \frac{x}{e^x-1}, \qquad x \in \mathbb{C}.$$

Useremo la notazione x, invece di s o z, per la variabile complessa poiché i risultati qui ottenuti verranno utilizzati nell'integrale all'interno della formula trovata alla fine del paragrafo precedente.
Questa funzione è analitica in $|x| < 2\pi$, poiché si hanno singolarità per $x = \pm 2\pi i$ (in quanto $e^{\pm 2\pi i} = 1$). Si potrebbe obiettare che anche $x = 0$ sia una singolarità dello stesso tipo della precedente, ma così non è e lo si può provare, ad esempio per x reale, con il teorema di L'Hôpital:

$$\lim_{x\to 0}\frac{x}{e^x-1} = \lim_{x\to 0}\frac{1}{e^x} = 1,$$

concetto si può estendere – con analoghe deduzioni – al caso di x complesso.
Considerando lo sviluppo in serie di potenze – o in serie di Taylor – della funzione $f(x)$ all'interno del disco $|x| < 2\pi i$ si trova che esso è della forma

$$\frac{x}{e^x-1} = \sum_{n=0}^{\infty}\frac{B_n x^n}{n!}.$$

I coefficienti B_n di tale sviluppo sono, per definizione, i numeri di Bernoulli (per $n \in \mathbb{N}$) [17].
Si provare che i numeri B_n, per n dispari, sono tutti nulli (tranne il primo): una dimostrazione semplice di questo fatto si otterrà con un procedimento a ritroso a partire dall'equazione funzionale della ζ. Vediamo, intanto, una tabella riassuntiva dei primi numeri di Bernoulli.

n	0	1	2	3	4	5	6	7	8
B_n	1	$-\frac{1}{2}$	$\frac{1}{6}$	0	$-\frac{1}{30}$	0	$\frac{1}{42}$	0	$-\frac{1}{30}$

I numeri di Bernoulli compaiono negli sviluppi di Taylor di varie funzioni (tra cui quella che li genera) ma, soprattutto – come vedremo in questo paragrafo – nella funzione ζ di Riemann.
Riprendiamo, a tal senso, la formula vista nel paragrafo precedente

$$\zeta(s) = \frac{\Gamma(1-s)}{2\pi i}\int_{+\infty}^{+\infty} \frac{(-x)^s}{x(e^x-1)}dx.$$

Vogliamo calcolare i valori della zeta per $s = -n$, con $n \in \mathbb{N}$, dunque per s intero negativo, applicando proprio lo sviluppo in termini contenenti i numeri di Bernoulli all'integrando. Tuttavia, prima di eseguire questo calcolo, occorre fare una precisazione.
Avevamo visto nel precedente paragrafo che

$$\int_{+\infty}^{+\infty} \frac{(-x)^s}{x(e^x-1)}dx = \left(e^{i\pi s} - e^{-i\pi s}\right)\int_0^{+\infty} \frac{x^{s-1}}{e^x-1}dx = 2i\sin(\pi s)\int_0^{+\infty} \frac{x^{s-1}}{e^x-1}dx.$$

Ponendo $s = -n$, con n naturale, l'intera espressione risulta essere uguale a zero (si annulla il seno). La parte dell'integrazione che resta da considerare è quella attorno all'origine che, per $\delta \to 0$ si annullava (nel calcolo precedente), però per $\delta > 0$ non è affatto nulla. Si può supporre $\delta = 1$, l'importante è che sia $0 < \delta < 2\pi$ per evitare le singolarità dell'integrando. Questo fatto giustifica il cambiamento di dominio di integrazione nel calcolo che ci accingiamo ad eseguire.

$$\begin{aligned}
\zeta(-n) &= \frac{\Gamma(1+n)}{2\pi i}\int_{|x|=\delta} \frac{(-x)^{-n}}{x(e^x-1)}dx = \frac{n!}{2\pi i}\int_{|x|=\delta} \frac{(-1)^{-n}x^{-n}}{x(e^x-1)}dx \\
&= \frac{n!\,(-1)^{-n}}{2\pi i}\int_{|x|=\delta} \frac{x^{-n}}{x(e^x-1)}dx = \frac{n!\,(-1)^{n}}{2\pi i}\int_{|x|=\delta} \left(\frac{x}{e^x-1}\right)\frac{x^{-n-1}}{x}dx \\
&= \frac{n!\,(-1)^{n}}{2\pi i}\int_{|x|=\delta} \left(\sum_{m=0}^{\infty}\frac{B_m x^m}{m!}\right)x^{-n-2}dx = \frac{n!\,(-1)^{n}}{2\pi i}\sum_{m=0}^{\infty}\frac{B_m}{m!}\int_{|x|=\delta} x^{m-n-2}dx \\
&= \frac{n!\,(-1)^{n}}{2\pi i}\sum_{m=0}^{\infty}\frac{B_m}{m!}\int_0^{2\pi} e^{it(m-n-2)}\cdot ie^{it}dt \\
&= \frac{n!\,(-1)^{n}}{2\pi i}\, i\sum_{m=0}^{\infty}\frac{B_m}{m!}\int_0^{2\pi} e^{it(m-n-1)}dt \\
&= \sum_{m=0}^{\infty}\frac{(-1)^n n!}{2\pi}\frac{B_m}{m!}\int_0^{2\pi}\big(\cos\big(t(m-n-1)\big) + i\sin\big(t(m-n-1)\big)\big)dt.
\end{aligned}$$

A questo punto abbiamo due casi, cioè $m = n+1$ e $m \neq n+1$.
Per $m = n+1$, vale

$$\int_0^{2\pi}\big(\cos\big(t(m-n-1)\big) + i\sin\big(t(m-n-1)\big)\big)dt = \int_0^{2\pi}\cos(0)\,dt = \int_0^{2\pi}dt = t|_0^{2\pi} = 2\pi.$$

Invece, per $m \neq n+1$, poniamo $k = m-n-1 \neq 0$, ottenendo

$$\int_0^{2\pi}\big(\cos\big(t(m-n-1)\big)+i\sin\big(t(m-n-1)\big)\big)dt$$
$$=\frac{\sin\big(t(m-n-1)\big)}{m-n-1}-i\frac{\cos\big(t(m-n-1)\big)}{m-n-1}\Bigg|_0^{2\pi}=\frac{\sin(kt)}{k}-i\frac{\cos(kt)}{k}\Bigg|_0^{2\pi}$$
$$=\frac{\sin(2k\pi)}{k}-\frac{i\cos(2k\pi)}{k}-\frac{\sin(0)}{k}+\frac{i\cos(0)}{k}=i\left(-\frac{1}{k}+\frac{1}{k}\right)=0.$$

Possiamo concludere che, nella sommatoria, si salva solo il termine $m=n+1$ (per il quale l'integrale vale 2π) poiché, per $m\neq n+1$, l'integrale è nullo e annulla tutto il corrispettivo termine della sommatoria stessa. Dunque

$$\sum_{m=0}^{\infty}\frac{(-1)^n n!}{2\pi}\frac{B_m}{m!}\int_0^{2\pi}\big(\cos\big(t(m-n-1)\big)+i\sin\big(t(m-n-1)\big)\big)dt=\frac{(-1)^n n!}{2\pi}\cdot\frac{2\pi B_{n+1}}{(n+1)!}$$
$$=(-1)^n\frac{B_{n+1}}{n+1}.$$

Quindi

$$\zeta(-n)=(-1)^n\frac{B_{n+1}}{n+1}.$$

La formula appena trovata ci consente di trarre alcune importanti conclusioni.

- $\zeta(-2n)=0$ (sempre con $n\in\mathbb{N}$): infatti corrisponde ad un numero di Bernoulli con indice dispari che sappiamo essere nullo. L'eccezione è $\zeta(0)=-B_1=-1/2$.
- Abbiamo una formula abbastanza agevole (in teoria) per il calcolo dei valori di $\zeta(s)$, con s intero negativo dispari.

Tuttavia, per quest'ultima affermazione, severi dubbi – poi confermati in certezze – spuntano fuori quando si cerca un algoritmo per il calcolo dei numeri di Bernoulli. Attualmente l'algoritmo più "efficiente" – se così si può dire – consiste proprio nell'applicare la formula della definizione e si impiega un numero di passi esponenziale rispetto all'input della ζ (segno a parte).

Possiamo comunque fornire qualche esempio pratico servendoci anche della tabella riassuntiva di alcuni numeri di Bernoulli vista all'inizio di questo paragrafo.

- $\zeta(-1)=(-1)^1\cdot\frac{1}{6}\cdot\frac{1}{2}=-\frac{1}{12}$,
- $\zeta(-3)=(-1)^3\cdot\left(-\frac{1}{30}\right)\cdot\frac{1}{4}=\frac{1}{120}$,

e così via.

Vale, inoltre, un simile risultato di Eulero che, tuttavia, non si può dedurre dalla formula integrale di Riemann per la ζ vista nel precedente paragrafo, ma lo dimostreremo a partire dall'equazione funzionale della ζ che vedremo nella prossima sottosezione:

$$\zeta(2n)=\frac{(2\pi)^{2n}(-1)^{n+1}B_{2n}}{2\cdot(2n)!}.$$

Tuttavia, nonostante queste interessanti formule, nulla si sa circa l'esistenza di una relazione che ci consenta di trovare $\zeta(2n+1)$ per n naturale, cioè i valori che assume la $\zeta(s)$ per s naturale dispari. Al giorno d'oggi si conoscono solo alcuni sviluppi particolari come, ad esempio:

- $\zeta(1)=\infty$ ($\zeta(1)$ corrisponde alla serie armonica semplice);
- $\zeta(3)=1{,}20205\ldots$ (costante di Apery ([13]))

e così via. Per alcuni valori interi positivi dispari della zeta si sono trovati degli sviluppi in serie piuttosto complicati ([12]) che, però, non traggono le loro fondamenta dall'equazione integrale vista nel precedente paragrafo.
Un esempio è proprio quella di Eulero per $\zeta(3)$ ([13])

$$\zeta(3) = \frac{\pi^2}{7}\left[1 - 4\sum_{k=1}^{\infty}\frac{\zeta(2k)}{(2k+1)(2k+2)2^{2k}}\right],$$

il cui valore fu in seguito trovato esplicitamente dal matematico Apery (da cui il nome).

12.3 EQUAZIONE FUNZIONALE PER LA ζ

In questo paragrafo introdurremo l'equazione funzionale per la funzione $\zeta(s)$. Si tratta di una formula che consente di esprimere i valori di $\zeta(1-s)$ tramite quelli di $\zeta(s)$: dunque, oltre a fornire un'estensione della ζ a tutto $\mathbb{C}\backslash\{1\}$, essa è anche un modo per calcolare i valori negativi della zeta servendosi di quelli positivi.
Il nome equazione funzionale, infatti, sta a indicare una relazione scritta in forma implicita, cioè del tipo

$$f(x) = h(x)f\big(g(x)\big),$$

in cui i valori di $f(x)$ si calcolano a partire da quelli di $f\big(g(x)\big)$ che si suppongono noti o per i quali sussiste un procedimento già noto per il calcolo. Nel caso della $\zeta(s)$, l'equazione funzionale è

$$\zeta(s) = h(s)\zeta(1-s).$$

La analizzeremo in quest'ultima sottosezione, specificando in particolare $h(s)$.
Ci occuperemo di dedurre questa equazione funzionale della ζ tramite le due diverse dimostrazioni date dallo stesso Riemann, in modo da rapportarci agevolmente al suo articolo di ricerca, riportato nell'appendice di questa tesi. In ([26], §2.1-2.10), si possono trovare – in maniera piuttosto sintetica – sette differenti dimostrazioni, tra cui le due di Riemann, circa l'equazione funzionale per la zeta.

12.3.1 Primo metodo di Riemann per l'equazione funzionale

Nel suo articolo di ricerca (§Appendice I), Riemann dà due dimostrazioni per l'equazione funzionale della funzione zeta. Questa è la prima di esse e si basa, per la maggior parte, sull'estensione della ζ a tutto il campo complesso (eccetto $s = 1$), vista in (§12.2.1):

$$\zeta(s) = \frac{\Gamma(1-s)}{2\pi i}\int_{+\infty}^{+\infty}\frac{(-x)^s}{x(e^x-1)}dx.$$

Analogamente alla dimostrazione precedente, anche alla base di questa c'è la valutazione dell'integrale lungo un cammino di integrazione alquanto inusuale. Tuttavia la strada è concettualmente più breve poiché ci si serve del teorema dei residui (§3.5.1).
La chiave sta nella corretta interpretazione di ciò che scrive Riemann nel suo articolo dopo la prima estensione analitica.

> <<Se la parte reale di s è negativa, allora, invece di essere preso in senso positivo intorno al dominio specifico, questo integrale può anche essere fatto in senso negativo intorno al dominio contenente tutte le restanti quantità complesse...>>
>
> (Tratto dall'articolo originale di Riemann)

Questa frase è la chiave di volta di tutto il discorso che faremo in questa sezione.
Analizziamola con calma.
Il punto di partenza è il cammino precedente, quello della Figura 12.1, fatto, però, in senso opposto: lo chiameremo C (Figura 12.3).

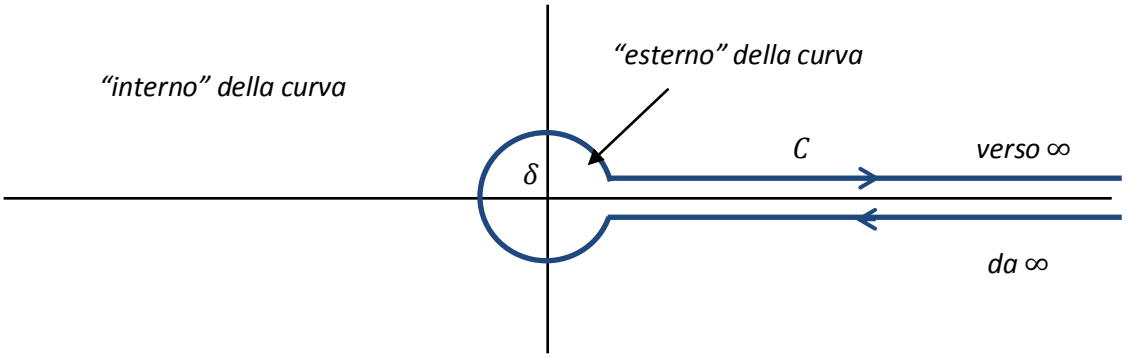

Figura 12.3. Nuovo contorno di integrazione.

In realtà, anche in questo caso l'integrale andrebbe inteso con la percorrenza lungo l'asse reale (nei due sensi) e non prima al di sotto poi al di sopra (Figura 12.4). Però, in molti testi si trova la Figura 12.3 perché riesce a focalizzare meglio la situazione e anche noi adotteremo questo metodo.

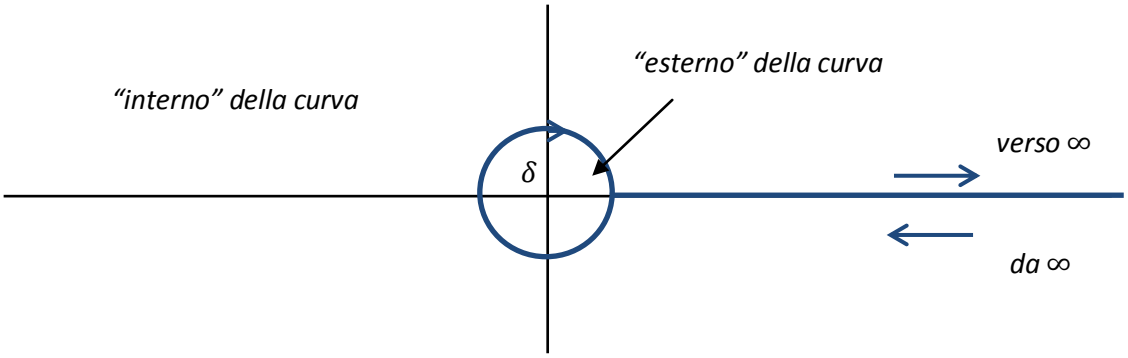

Figura 12.4. Nuovo contorno di integrazione.

La questione di come influisce il verso di percorrenza andrebbe approfondita più nel dettaglio, tuttavia servirebbero questioni tecniche di topologia che, in questa tesi, abbiamo omesso perché non inerenti allo scopo della stessa. Tuttavia diamo per buono il fatto che, cambiando verso di percorrenza, quello che prima era l'interno ora è l'esterno e viceversa.

Ora Riemann passa direttamente alle conclusioni applicando il teorema dei residui (non in senso classico), dicendo che "l'integrando ha discontinuità solo se x diventa uguale a tutti i multipli di $\pm 2\pi i$, e l'integrale è così uguale alla somma degli integrali fatti in senso negativo intorno a questi valori" (§Appendice I).
Questa sarà una conclusione più che lecita, la quale, tuttavia, merita ulteriori approfondimenti. Rispetto a prima, però, la curva non è chiusa, dunque non possiamo applicare il teorema dei residui senza i dovuti accorgimenti: la differenza con il cammino precedente sta nel fatto che l'interno di questo è l'esterno del precedente e viceversa a causa del verso opposto di percorrenza.
Stavolta all'interno ci sono i poli dell'integrando (Figura 12.5).

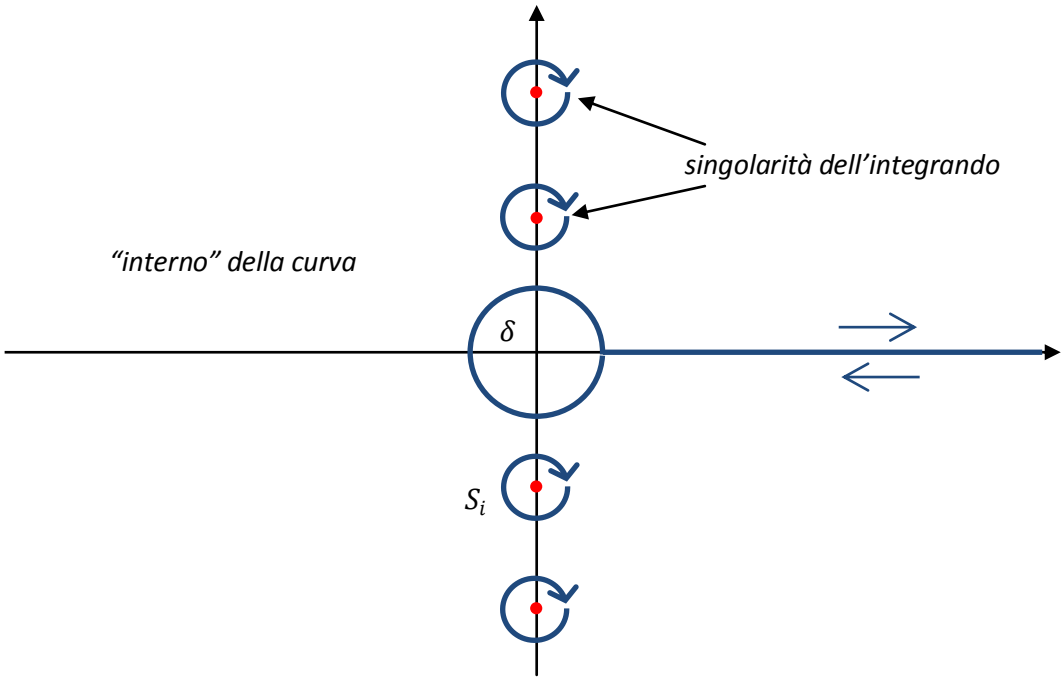

Figura 12.5. Stavolta l'interno della curva di integrazione contiene le singolarità dell'integrando (punti $2n\pi i$, in rosso). Le circonferenze in senso antiorario attorno a questi sono "le circonferenze in senso negativo" di cui parla Riemann e vengono indicate con S_i.

All'interno del cammino ci sono le singolarità dell'integrando: vorremmo, dunque, rapportarci al teorema dei residui (§3.5.1) ricordando che l'indice di avvolgimento è 1 (§3.3.3)

$$\int_{+\infty}^{+\infty} \frac{(-x)^s}{x(e^x-1)}dx = \int_{\gamma} \frac{(-x)^s}{x(e^x-1)}dx = 2\pi i \sum_{i=1}^{n} Res\left(\frac{(-x)^s}{x(e^x-1)}, \pm 2m\pi i\right) Ind_{\gamma}(\pm 2m\pi i),$$

ma questo non vale perché la curva non è chiusa e il suo interno è illimitato. Per ovviare a questo dettaglio tecnico si utilizza la nozione di limite nel modo indicato in Figura 12.6.

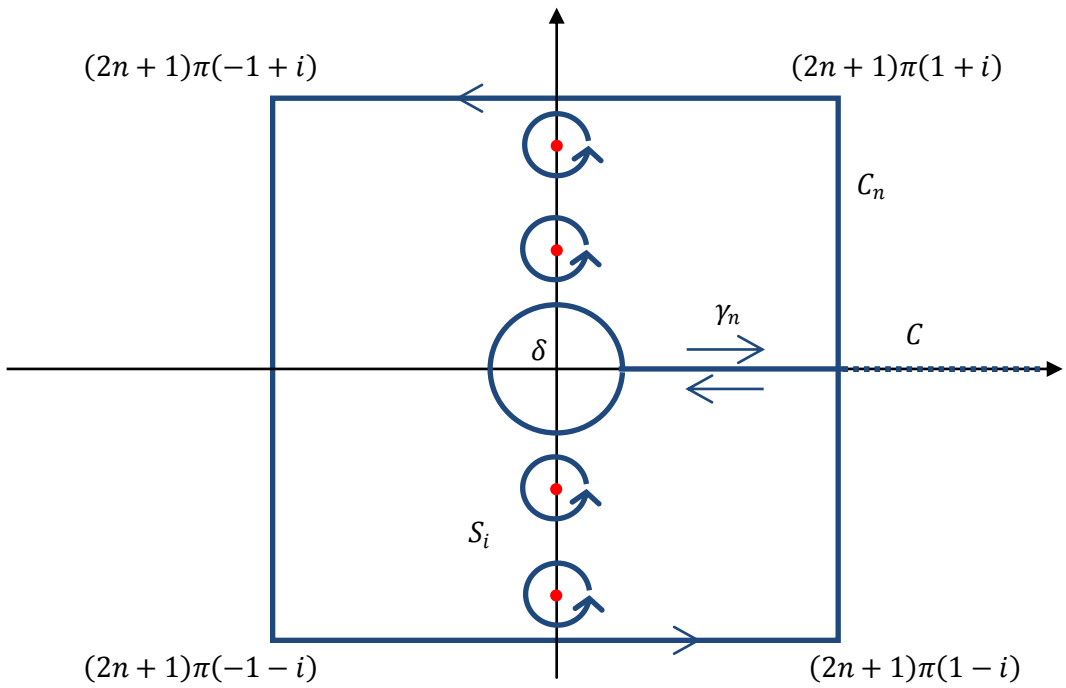

Figura 12.6. Costruzione del quadrato C_n attorno al cammino C di integrazione.

Il quadrato C_n ha come vertici i quattro punti del piano complesso indicati, sinteticamente, con $(2n+1)\pi(\pm 1 \pm i)$ (si veda la Figura 12.6). Nel complesso, questo nuovo cammino lo chiameremo $C_n + \gamma_n$ a sottolineare come i vertici del quadrato e γ_n varino con n.
Descriviamo $C_n + \gamma_n$:

- γ_n è composta da un tratto sull'asse reale (percorso in doppio senso) e dalla circonferenza $|x| = \delta$;
- C_n è il quadrato di vertici $(2n+1)\pi(\pm 1 \pm i)$.

Quindi, il cammino si percorre nel seguente modo (fare fede alla Figura 12.6).
A partire dall'asse reale si percorre verso sinistra il segmento di γ_n (che giace sull'asse reale) fino ad arrivare alla circonferenza $|x| = \delta$ e la si percorre in senso antiorario. In seguito si torna nuovamente sull'asse reale percorrendo il segmento orizzontale di γ_n verso destra fino ad incontrare il quadrato percorrendolo in senso antiorario per poi tornare nuovamente in γ_n.
Possiamo notare che, per $n \to \infty, \gamma_n \to C$, in cui C è il cammino di partenza descritto ad inizio paragrafo.
Il risultato è una curva chiusa ed è in questo cammino che dobbiamo utilizzare il teorema dei residui, notando che ci sono delle singolarità all'interno della regione contenuta in quella curva.
Per la sua complessità, ometteremo il calcolo del residuo nei punti nei punti $2mi\pi$ e $-2mi\pi$ per $1 \leq m \leq n$ – cioè i poli contenuti nella zona tra C e C_n – che vale

$$-2(2m\pi)^{s-1}e^{i\pi(s-1)}\sin\left(\frac{1}{2}\pi s\right).$$

Per il teorema dei residui (l'indice di avvolgimento intorno a ciascuno di essi è 1)

$$\int_{C_n+\gamma_n} \frac{(-x)^s}{x(e^x-1)}dx = 2\pi i \sum_{m=1}^{n} Res\left(\frac{(-x)^s}{x(e^x-1)}dx, \pm 2m\pi i\right)$$
$$= 2\pi i \sum_{m=1}^{n} \left(-2(2m\pi)^{s-1}e^{i\pi(s-1)}\sin\left(\frac{1}{2}\pi s\right)\right).$$

Facciamo, dunque, tendere $n \to \infty$ ricordando che dobbiamo calcolare l'integrale su C, cioè su γ_n (per $n \to \infty, \gamma_n \to C$). Ricordiamo, inoltre, che per gli integrali su curve regolari a tratti

$$\int_{C_n+\gamma_n} \frac{(-x)^s}{x(e^x-1)}dx = \int_{C_n} \frac{(-x)^s}{x(e^x-1)}dx + \int_{\gamma_n} \frac{(-x)^s}{x(e^x-1)}dx$$

e, nel nostro caso,

$$\int_{\gamma_n} \frac{(-x)^s}{x(e^x-1)}dx = \int_{C_n+\gamma_n} \frac{(-x)^s}{x(e^x-1)}dx - \int_{C_n} \frac{(-x)^s}{x(e^x-1)}dx.$$

Prima di passare al calcolo esplicito, notiamo che l'integrale su C_n tende a zero per $n \to \infty$. Premettiamo che, in molti testi, l'integrale fatto su C_n ha segno positivo: il motivo sta nel fatto che C_n (che, come detto, a volte è inteso come cerchio) è percorso in senso orario e quindi il verso opposto annulla il meno davanti all'integrale stesso (un esempio sta in ([28], §6.2), in esso il cammino è proprio opposto a quello indicato da Riemann anche se le conclusioni sono le medesime).

Tornando all'integrale su C_n, scrivendolo come

$$\int_{C_n} \frac{(-x)^s}{x(e^x-1)}dx = \int_{C_n} \frac{(-x)^{s-1}}{e^x-1}dx,$$

si ha:

- $\frac{1}{e^x-1}$ è una funzione limitata lungo C_n (in quanto è analitica e non ha poli);
- per x^{s-1}, invece, sapendo che s è negativo

$$|x^{s-1}| = |x|^{s-1} = n^{s-1} \xrightarrow{n\to\infty} 0.$$

Da questo otteniamo che l'integrando tende a zero per $n \to \infty$ poiché è il risultato del prodotto tra un termine costante e uno infinitesimo. Per il teorema di assoluta continuità dell'integrale concludiamo che l'integrale stesso è nullo.

Procediamo, dunque, con il calcolo esplicito tenendo conto delle proprietà appena ricordate.

$$\int_{+\infty}^{+\infty} \frac{(-x)^s}{x(e^x-1)}dx = \lim_{n\to\infty}\left(2\pi i \sum_{m=1}^{n}\left(-2(2m\pi)^{s-1}e^{i\pi(s-1)}\sin\left(\frac{1}{2}\pi s\right)\right)\right)$$
$$= 2\pi i \cdot (-2) \cdot e^{i\pi(s-1)} \cdot \sin\left(\frac{1}{2}\pi s\right) \cdot \sum_{m=1}^{\infty}(2m\pi)^{s-1}$$
$$= -4\pi i e^{i\pi s} e^{-i\pi} \sin\left(\frac{1}{2}\pi s\right) \sum_{m=1}^{\infty}(2m\pi)^{s-1}$$
$$= 4\pi i e^{i\pi s} \sin\left(\frac{1}{2}\pi s\right)(2\pi)^{s-1} \sum_{m=1}^{\infty} m^{s-1}$$
$$= 4\pi i e^{i\pi s} \sin\left(\frac{1}{2}\pi s\right)(2\pi)^{s-1}\zeta(1-s),$$

quest'ultimo passaggio giustificato dalla definizione stessa della ζ di Riemann

$$\zeta(1-s) = \sum_{m=1}^{\infty} m^{s-1} = \sum_{m=1}^{\infty} \frac{1}{m^{1-s}},$$

ricordando che questo discorso è ben posto poiché s è negativo, dunque $1-s>0$.
Otteniamo, dunque

$$\int_{+\infty}^{+\infty} \frac{(-x)^s}{x(e^x-1)} dx = 4\pi i e^{i\pi s} \sin\left(\frac{1}{2}\pi s\right)(2\pi)^{s-1}\zeta(1-s).$$

Ricordando l'estensione della $\zeta(s)$ operata in (§12.2.1), concludiamo

$$\begin{aligned}\zeta(s) &= \frac{\Gamma(1-s)}{2\pi i}\int_{+\infty}^{+\infty} \frac{(-x)^s}{x(e^x-1)} dx = \frac{\Gamma(1-s)}{2\pi i} 4\pi i e^{i\pi s} \sin\left(\frac{1}{2}\pi s\right)(2\pi)^{s-1}\zeta(1-s)\\ &= \Gamma(1-s)2e^{i\pi s}\sin\left(\frac{1}{2}\pi s\right)(2\pi)^{s-1}\zeta(1-s)\\ &= 2\Gamma(1-s)(2\pi)^{s-1}\sin\left(\frac{\pi s}{2}\right)\zeta(1-s).\end{aligned}$$

Cioè

$$\zeta(s) = 2\Gamma(1-s)(2\pi)^{s-1}\sin\left(\frac{\pi s}{2}\right)\zeta(1-s),$$

che è l'equazione funzionale per la $\zeta(s)$ con s negativo.

12.3.2 Osservazioni importanti dall'equazione funzionale

Presentiamo alcune considerazioni riferite all'equazione funzionale trovata nel paragrafo precedente.

(i) L'equazione funzionale dimostra inequivocabilmente che l'estensione analitica della ζ di Riemann a tutto il piano complesso – eccetto il punto $s=1$ – è, a conti fatti, analitica, in quanto i termini al secondo membro sono prodotti di funzioni analitiche.
Ricordiamo, inoltre, che s è tale che $Re(s)<0$, dunque al secondo membro $\zeta(1-s)$ è definita nel modo classico poiché $Re(1-s)>1$.

(ii) Questa equazione è valida per $Re(s)<0$: tutte le funzioni al secondo membro sono analitiche.

(iii) Per $s=-2n$ con n intero positivo, otteniamo

$$\zeta(-2n) = 2\Gamma\big(1-(-2n)\big)(2\pi)^{-2n-1}\sin\left(\frac{-2n\pi}{2}\right)\zeta\big(1-(-2n)\big),$$

che, a prescindere dal fatto che $\Gamma(2n+1)=(2n)!$ o altre semplificazioni, ci mostra che $\zeta(-2n)=0$ poiché si annulla il seno. Tutto questo è valido perché le altre quantità sono tutte limitate a cominciare da $\zeta\big(1-(-2n)\big)=\zeta(2n+1)<\zeta(2)$ per definizione della ζ come generalizzazione della serie armonica e per il fatto che n è intero positivo.
Questi zeri della funzione ζ sono detta "zeri banali": il termine sta proprio a sottolineare che sono dovuti ad un "banale" annullamento della funzione seno che compare nell'equazione funzionale.

(iv) Ricordiamo la formula vista in precedenza, indici a parte, per il calcolo dei valori interi negativi della zeta (§12.2.2)

$$\zeta(-k) = (-1)^k \frac{B_{k+1}}{k+1};$$

ponendo $k = 2n - 1$, otteniamo

$$\zeta(1-2n) = (-1)^{2n-1} \frac{B_{2n}}{2n}.$$

Ora, poiché $(-1)^{2n-1} = -1$ per qualsiasi scelta di n intero, abbiamo

$$\zeta(1-2n) = -\frac{B_{2n}}{2n}$$

e dunque

$$B_{2n} = -2n\zeta(1-2n).$$

A questo punto possiamo applicare l'equazione funzionale della funzione ζ, ottenendo

$$\zeta(1-2n) = 2\Gamma\big(1-(1-2n)\big)(2\pi)^{1-2n-1} \sin\left(\frac{\pi(1-2n)}{2}\right)\zeta\big(1-(1-2n)\big)$$
$$= 2\Gamma(2n)(2\pi)^{-2n} \sin\left(\frac{\pi(1-2n)}{2}\right)\zeta(2n).$$

Prima di andare avanti possiamo notare che

$$\sin\left(\frac{\pi(1-2n)}{2}\right) = (-1)^n,$$

poiché, al variare di n naturale, assume solo valore -1 o 1: il primo per n dispari mentre il secondo per n pari.

Adesso, ponendo all'interno di questa espressione la relazione $B_{2n} = -2n\zeta(1-2n)$ trovata in precedenza, otteniamo

$$B_{2n} = -2n\Gamma(2n)(2\pi)^{-2n} 2 \sin\left(\frac{\pi(1-2n)}{2}\right)\zeta(2n)$$
$$= -(2n)!\,(2\pi)^{-2n}(-1)^n\zeta(2n) = (2n)!\,(2\pi)^{-2n}(-1)^{n+1}\zeta(2n)$$

Dunque

$$\zeta(2n) = \frac{B_{2n}(2\pi)^{2n}(-1)^{n+1}}{2(2n)!},$$

che è la formula trovata da Eulero (anni prima di Riemann e dunque non in questo modo!) e accennata in precedenza.

(v) Ricordando la formula di duplicazione per la funzione Gamma (§8.2)

$$\Gamma(2s) = \frac{2^{2s-1}}{\sqrt{\pi}}\Gamma(s)\Gamma\left(s+\frac{1}{2}\right),$$

applichiamola, insieme alla formula di riflessione (§8.3), all'equazione funzionale della ζ:

$$\zeta(s) = 2\Gamma(1-s)(2\pi)^{s-1} \sin\left(\frac{\pi s}{2}\right)\zeta(1-s).$$

In essa sostituiremo $\Gamma(1-s)$ con la formula di duplicazione appena richiamata e il $\sin\left(\frac{\pi s}{2}\right)$ con la formula di riflessione. Ovviamente esse vanno opportunamente sistemate e adattate:

$$\Gamma(1-s) = \frac{2^{(1-s)-1}}{\sqrt{\pi}}\Gamma\left(\frac{1-s}{2}\right)\Gamma\left(\frac{1-s}{2}+\frac{1}{2}\right) = 2^{-s}\pi^{-\frac{1}{2}}\Gamma\left(\frac{1-s}{2}\right)\Gamma\left(1-\frac{s}{2}\right),$$

per la prima e

$$\sin\left(\frac{\pi s}{2}\right) = \frac{\pi}{\Gamma(s/2)\Gamma(1-s/2)},$$

per la seconda. Andiamo, dunque, a sostituire

$$\begin{aligned}\zeta(s) &= 2\Gamma(1-s)(2\pi)^{s-1}\sin\left(\frac{\pi s}{2}\right)\zeta(1-s)\\ &= \left[2^{-s}\pi^{-\frac{1}{2}}\Gamma\left(\frac{1-s}{2}\right)\Gamma\left(1-\frac{s}{2}\right)\right]\cdot 2\cdot 2^{s-1}\cdot\pi^{s-1}\\ &\cdot\left[\frac{\pi}{\Gamma(s/2)\Gamma(1-s/2)}\right]\cdot\zeta(1-s)\\ &= 2^{-s}\pi^{-\frac{1}{2}}\Gamma\left(\frac{1-s}{2}\right)\Gamma\left(1-\frac{s}{2}\right)2^{s}\pi^{s}\frac{1}{\Gamma(s/2)\Gamma(1-s/2)}.\end{aligned}$$

A questo punto semplifichiamo, nella catena di moltiplicazioni, i termini $\Gamma(1-s/2)$, e 2^{-s} per ottenere

$$\zeta(s) = \left[\pi^{-\frac{1}{2}}\Gamma\left(\frac{1-s}{2}\right)\right]\cdot\pi^{s}\cdot\left[\frac{1}{\Gamma(s/2)}\right]\zeta(1-s),$$

cioè

$$\Gamma\left(\frac{s}{2}\right)\zeta(s) = \pi^{\frac{s}{2}+\frac{s}{2}-\frac{1}{2}}\Gamma\left(\frac{1-s}{2}\right)\zeta(1-s),$$

quindi

$$\Gamma\left(\frac{s}{2}\right)\zeta(s)\pi^{-\frac{s}{2}} = \Gamma\left(\frac{1-s}{2}\right)\pi^{\frac{s-1}{2}}\zeta(1-s) = \Gamma\left(\frac{1-s}{2}\right)\pi^{-\frac{1-s}{2}}\zeta(1-s),$$

concludendo

$$\Gamma\left(\frac{s}{2}\right)\zeta(s)\pi^{-\frac{s}{2}} = \Gamma\left(\frac{1-s}{2}\right)\pi^{-\frac{1-s}{2}}\zeta(1-s).$$

In questa formula è evidente una simmetria nel passaggio da s a $1-s$. Se al primo membro si sostituisce $1-s$ a s si ottiene il secondo membro e viceversa per il secondo: in altre parole c'è simmetria lungo l'asse $Re(s) = \frac{1}{2}$.

Infatti, si può verificare che se in quest'ultima equazione si pone $s = \frac{1}{2}$ (anche se vale, in generale, per $Re(s) = \frac{1}{2}$, di cui $s = \frac{1}{2}$ è un caso particolare) al primo membro si ottiene

$$\Gamma\left(\frac{s}{2}\right)\zeta(s)\pi^{-\frac{s}{2}}\Big|_{s=\frac{1}{2}} = \Gamma\left(\frac{1}{4}\right)\zeta\left(\frac{1}{2}\right)\pi^{-\frac{1}{4}},$$

mentre al secondo

$$\begin{aligned}\Gamma\left(\frac{1-s}{2}\right)\pi^{-\frac{1-s}{2}}\zeta(1-s)\Big|_{s=\frac{1}{2}} &= \Gamma\left(\frac{1-1/2}{2}\right)\pi^{-\frac{1-1/2}{2}}\zeta\left(1-\frac{1}{2}\right)\\ &= \Gamma\left(\frac{1}{4}\right)\pi^{-\frac{1}{4}}\zeta\left(\frac{1}{2}\right),\end{aligned}$$

che è lo stesso risultato al primo membro. Questo dimostra anche che non si può utilizzare l'equazione funzionale per il calcolo di $\zeta\left(\frac{1}{2}\right)$ in quanto i termini finirebbero per semplificarsi reciprocamente.

Tuttavia, questo non vuol dire affatto che $\zeta(s)$ – cioè la ζ di Riemann più la sua estensione analitica – non è definita per $Re(s) = \frac{1}{2}$, ma solo che, per calcolare

$$\zeta\left(\frac{1}{2}+it\right),$$

cioè i valori di $\zeta(s)$ per $Re(s) = \frac{1}{2}$ ci si deve servire delle altre formule, per esempio quella integrale (§12.2.1) o quella ottenuta tramite la η di Dirichlet (12.1.2).

Riprenderemo il discorso sulla simmetria nella prossima sezione, parlando della funzione ξ di Riemann definita per mezzo dell'equazione funzionale della ζ.

(vi) Un'ultima osservazione.

Si era calcolata l'equazione funzionale con il metodo di Riemann, ottenendo

$$\zeta(s) = 2\Gamma(1-s)(2\pi)^{s-1}\sin\left(\frac{\pi s}{2}\right)\zeta(1-s);$$

se poniamo $s = 1 - z$, con $z \in \mathbb{C}$, otteniamo

$$\zeta(1-z) = 2\Gamma(z)(2\pi)^{-z}\sin\left(\frac{\pi(1-z)}{2}\right)\zeta(z),$$

che è l'equazione funzionale così come compare in molti testi trattanti l'ipotesi di Riemann. La sua logica è quella di calcolare $\zeta(1-z)$ a partire da $\zeta(z)$ che è un discorso analogo – anche se da un punto di vista differente – rispetto a quello visto in precedenza.
Si può anche sostituire nuovamente z con s per quanto riguarda la variabile complessa che compare nell'equazione, ma è solo una questione di dettagli estetici.

12.3.3 Le funzioni θ e ψ di Jacobi

Prima di mostrare la seconda prova di Riemann dell'equazione funzionale, occorre definire due funzioni intermedie: le funzioni θ e ψ di Jacobi.
La notazione potrebbe far confondere, in quanto con la lettera greca ψ si indica un'altra funzione di Chebyshev, tuttavia solo in questo (e in pochi altri paragrafi) utilizzeremo le due funzioni di Jacobi come passi intermedi per mostrare la seconda prova di Riemann dell'equazione funzionale per la ζ. Per quanto riguarda la θ, invece, c'è meno confusione poiché la funzione di Chebyshev è indicata con una variante della theta stessa, cioè ϑ (§10.2.4).
Definiamo, dunque, le due funzioni di Jacobi. Per $x > 0$ reale

$$\theta(x) = \sum_{n=-\infty}^{\infty} e^{-n^2\pi x}$$

è la funzione θ di Jacobi, mentre

$$\psi(x) = \sum_{n=1}^{\infty} e^{-n^2\pi x}$$

è la funzione ψ di Jacobi (quella da non confondere con la ψ di Chebyshev).
Si può notare che vale la seguente relazione

$$\theta(x) = 2\psi(x) + 1,$$

infatti l'indice n degli esponenti compare sempre al quadrato, dunque, con abuso di scrittura,

$$\sum_{n=-\infty}^{-1} e^{-n^2\pi x} = \sum_{n=1}^{\infty} e^{-n^2\pi x} = \psi(x).$$

Inoltre l'addendo finale "$+1$" deriva dall'aggiunta del termine corrispondente all'indice $n = 0$ che vale, proprio, $e^{-n^2\pi x} = e^0 = 1$.
La proprietà più importante per queste due funzioni è deducibile direttamente dalla loro definizione. Prendiamo, ad esempio, $\psi(x)$,

$$\psi(x) = \sum_{n=1}^{\infty} e^{-n^2\pi x},$$

per ogni $x > 0$ fissato, la serie converge totalmente in quanto è sempre possibile trovare un esponenziale per cui maggiorarla termine a termine. La convergenza, dunque, è anche assoluta e uniforme e quindi ψ è continua. Tuttavia, ciò vale anche per θ grazie alla relazione

$$\theta(x) = 2\psi(x) + 1.$$

Per queste due funzioni vale anche un'altra proprietà molto interessante, la cui dimostrazione sarà omessa poiché si serve di metodi sofisticati riguardanti la trattazione delle funzioni in analisi complessa che non abbiamo richiamato poiché non inerenti ai fini della comprensione dell'ipotesi di Riemann.

Teorema (identità di Jacobi)
Per $x > 0$, si ha

$$\sum_{n=-\infty}^{\infty} e^{-n^2\pi x} = \frac{1}{\sqrt{x}} \sum_{n=-\infty}^{\infty} e^{-\frac{n^2\pi}{x}}.$$

In riferimento a θ questo equivale a dire che

$$\theta(x) = \frac{1}{\sqrt{x}}\theta\left(\frac{1}{x}\right),$$

mentre per quanto riguarda ψ

$$2\psi(x) + 1 = \frac{1}{\sqrt{x}}\left(2\psi\left(\frac{1}{x}\right) + 1\right).$$

Quest'ultima equazione è anche chiamata equazione funzionale della ψ, infatti, isolando il termine $\psi(x)$ al primo membro, si ottiene

$$\psi(x) = \frac{1}{\sqrt{x}}\psi\left(\frac{1}{x}\right) + \frac{1}{2\sqrt{x}} - \frac{1}{2}.$$

A questo punto possiamo procedere alla seconda dimostrazione di Riemann per l'equazione funzionale.

12.3.4 Secondo metodo utilizzato da Riemann

Vediamo, dunque, un secondo metodo usato da Riemann (§Appendice I) per mostrare l'equazione funzionale della ζ. E' una procedura molto meccanica che consiste nell'utilizzare le proprietà delle funzioni di Jacobi per la valutazione di alcuni integrali.
Partiamo dalla definizione della funzione Γ (§8)

$$\Gamma(s) = \int_0^{\infty} t^{s-1}e^{-t}dt,$$

scrivendola per $\frac{s}{2}$ invece che per s

$$\Gamma\left(\frac{s}{2}\right) = \int_0^{\infty} t^{\frac{s}{2}-1}e^{-t}\,dt.$$

A questo punto cambiamo variabile, ponendo $t = n^2\pi x$ e quindi $dt = n^2\pi dx$ nell'integrale.

$$\Gamma\left(\frac{s}{2}\right) = \int_0^\infty (n^2\pi x)^{\frac{s}{2}-1} e^{-n^2\pi x} n^2\pi dx = \int_0^\infty n^{s-2}\pi^{\frac{s}{2}-1} x^{\frac{s}{2}-1} e^{-n^2\pi x} n^2\pi dx$$

$$= \int_0^\infty n^s \pi^{\frac{s}{2}} x^{\frac{s}{2}-1} e^{-n^2\pi x}\, dx = n^s\pi^{\frac{s}{2}} \int_0^\infty x^{\frac{s}{2}-1} e^{-n^2\pi x}\, dx.$$

Portando le costanti che non dipendono dalla variabile d'integrazione al primo membro, otteniamo

$$n^{-s}\pi^{-\frac{s}{2}}\Gamma\left(\frac{s}{2}\right) = \int_0^\infty x^{\frac{s}{2}-1} e^{-n^2\pi x}\, dx.$$

Ora possiamo sommare ambo i membri per n da uno a infinito

$$\sum_{n=1}^\infty n^{-s}\, \pi^{\frac{s}{2}}\Gamma\left(\frac{s}{2}\right) = \sum_{n=1}^\infty \int_0^\infty x^{\frac{s}{2}-1} e^{-n^2\pi x}\, dx.$$

Il punto sta nell'isolare i termini che dipendono dall'indice di sommatoria rispetto agli altri. Esaminiamo, infatti, separatamente i due membri dell'equazione ottenuta.

$$\sum_{n=1}^\infty n^{-s}\, \pi^{\frac{s}{2}}\Gamma\left(\frac{s}{2}\right) = \pi^{\frac{s}{2}}\Gamma\left(\frac{s}{2}\right)\sum_{n=1}^\infty n^{-s} = \pi^{\frac{s}{2}}\Gamma\left(\frac{s}{2}\right)\zeta(s)$$

per quanto riguarda il primo membro, mentre

$$\sum_{n=1}^\infty \int_0^\infty x^{\frac{s}{2}-1} e^{-n^2\pi x}\, dx = \int_0^\infty x^{\frac{s}{2}-1} \sum_{n=1}^\infty e^{-n^2\pi x}\, dx = \int_0^\infty x^{\frac{s}{2}-1}\psi(x)dx$$

al secondo membro. In quest'ultimo abbiamo utilizzato, seppur in senso inverso, la seguente proprietà delle serie di funzioni (§2.2.2)

$$\int_a^b \sum_n f_n(x)\, dx = \sum_n \int_a^b f_n(x)dx,$$

poiché la serie che ha come somma $\psi(x)$ converge uniformemente, come si è visto nel paragrafo precedente. Riunendo i due membri, otteniamo

$$\pi^{\frac{s}{2}}\Gamma\left(\frac{s}{2}\right)\zeta(s) = \int_0^\infty x^{\frac{s}{2}-1}\psi(x)dx = \int_0^1 x^{\frac{s}{2}-1}\psi(x)dx + \int_1^\infty x^{\frac{s}{2}-1}\psi(x)dx.$$

Si tratta di utilizzare l'identità di Jacobi nel primo dei due integrali, ricordandosi di sfruttare la linearità dell'integrale stesso.

$$\begin{aligned}
\int_0^1 x^{\frac{s}{2}-1}\psi(x)dx &= \int_0^1 \left(\frac{x^{\frac{s}{2}-1}}{\sqrt{x}}\psi\left(\frac{1}{x}\right) + \frac{x^{\frac{s}{2}-1}}{2\sqrt{x}} - \frac{x^{\frac{s}{2}-1}}{2}\right)dx \\
&= \int_0^1 \left(x^{\frac{s}{2}-1-\frac{1}{2}}\psi\left(\frac{1}{x}\right) + \frac{1}{2}x^{\frac{s}{2}-1-\frac{1}{2}} - \frac{1}{2}x^{\frac{s}{2}-1}\right)dx \\
&= \int_0^1 x^{\frac{s}{2}-\frac{3}{2}}\psi\left(\frac{1}{x}\right)dx + \int_0^1 \frac{1}{2}x^{\frac{s}{2}-\frac{3}{2}}dx - \int_0^1 \frac{1}{2}x^{\frac{s}{2}-1}dx \\
&= \int_0^1 x^{\frac{s}{2}-\frac{3}{2}}\psi\left(\frac{1}{x}\right)dx + \left.\frac{x^{\frac{s}{2}-\frac{1}{2}}}{2\left(\frac{s}{2}-\frac{1}{2}\right)}\right|_0^1 - \left.\frac{x^{\frac{s}{2}}}{2\left(\frac{s}{2}\right)}\right|_0^1 \\
&= \int_0^1 x^{\frac{s}{2}-\frac{3}{2}}\psi\left(\frac{1}{x}\right)dx + \left(\frac{1}{s-1} - 0\right) - \left(\frac{1}{s} - 0\right) \\
&= \int_0^1 x^{\frac{s}{2}-\frac{3}{2}}\psi\left(\frac{1}{x}\right)dx + \frac{1}{s-1} - \frac{1}{s} = \int_\infty^1 y^{-\frac{s}{2}+\frac{3}{2}}\psi(y)(-y^2)dx + \frac{s-s+1}{s(s-1)} \\
&= -\int_\infty^1 y^{-\frac{s}{2}-\frac{1}{2}}\psi(y)dy + \frac{1}{s(s-1)} = \int_1^\infty y^{-\frac{s}{2}-\frac{1}{2}}\psi(y)dy + \frac{1}{s(s-1)}.
\end{aligned}$$

Nel corso di questo calcolo si era posto $y = \frac{1}{x}$, con y che varia da ∞ a 1 poiché x variava da 0 a 1 e $dx = -\frac{1}{y^2}dy$. Possiamo, dunque, tornare al calcolo principale ricordandosi di identificare nuovamente y con x (ovviamente la x non è la stessa di prima, si tratta di una formalità).

$$\begin{aligned}
\pi^{\frac{s}{2}}\Gamma\left(\frac{s}{2}\right)\zeta(s) &= \int_0^\infty x^{\frac{s}{2}-1}\psi(x)dx = \int_0^1 x^{\frac{s}{2}-1}\psi(x)dx + \int_1^\infty x^{\frac{s}{2}-1}\psi(x)dx \\
&= \int_1^\infty y^{-\frac{s}{2}-\frac{1}{2}}\psi(y)dy + \int_1^\infty x^{\frac{s}{2}-1}\psi(x)dx + \frac{1}{s(s-1)} \\
&= \int_1^\infty x^{-\frac{s}{2}-\frac{1}{2}}\psi(x)dx + \int_1^\infty x^{\frac{s}{2}-1}\psi(x)dx + \frac{1}{s(s-1)} \\
&= \int_1^\infty \left(x^{-\frac{s}{2}-\frac{1}{2}} + x^{\frac{s}{2}-1}\right)\psi(x)dx + \frac{1}{s(s-1)}.
\end{aligned}$$

Poiché l'integrale converge $\forall x$ in quanto la $\psi(x)$, come somma di esponenziali (a esponente negativo) domina una qualsiasi potenza di x, abbiamo l'ennesima dimostrazione dell'analiticità della ζ in tutto $\mathbb{C}$.

Nel suo articolo (§Appendice I), Riemann conclude con questo risultato per poi passare oltre, definendo la funzione ξ che vedremo nella prossima sezione. Infatti, questa dimostrazione si conclude osservando che, se nel secondo membro dell'identità trovata, cioè

$$\pi^{\frac{s}{2}}\Gamma\left(\frac{s}{2}\right)\zeta(s) = \int_1^\infty \left(x^{-\frac{s}{2}-\frac{1}{2}} + x^{\frac{s}{2}-1}\right)\psi(x)dx + \frac{1}{s(s-1)},$$

si sostituisce $1-s$ a s, si ottiene

$$\pi^{\frac{1-s}{2}}\Gamma\left(\frac{1-s}{2}\right)\zeta(1-s) = \int_1^\infty \left(x^{-\frac{s}{2}-\frac{1}{2}} + x^{\frac{s}{2}-1}\right)\psi(x)dx + \frac{1}{s(s-1)},$$

ovvero che il secondo membro non cambia sostituendo s con $1-s$. Si può dunque concludere, per la proprietà transitiva

$$\pi^{\frac{s}{2}}\Gamma\left(\frac{s}{2}\right)\zeta(s) = \pi^{\frac{1-s}{2}}\Gamma\left(\frac{1-s}{2}\right)\zeta(1-s),$$

poiché entrambi i membri sono uguali allo stesso termine.

12.3.5 Altri metodi per la prova dell'equazione funzionale

Generalmente nei testi riguardanti la trattazione della ζ o dell'ipotesi di Riemann, si citano solo le due dimostrazioni (a volte anche una sola delle due) fatte dal matematico tedesco per l'equazione funzionale della ζ, le stesse viste in questa sezione.
Titchmarsh ([26], §2.1-2.10) mostra ben sette differenti modi per la ricerca dell'equazione funzionale ognuno basato su una strategia operativa differente.
Tuttavia, quelli più "semplici" restano i metodi utilizzati da Riemann, anche se il matematico tedesco non ha spiegato le procedure nel dettaglio, lasciandole "intuire" dagli addetti ai lavori.
Nel già citato testo di Titchmarsh ([26]), esse costituiscono, rispettivamente, il "secondo" e il "terzo" metodo (tra i sette mostrati) per la dimostrazione dell'equazione funzionale della ζ.
L'autore, infatti, ha cura di aprire il paragrafo con un "this is one of Riemann's original proof" (= "è una delle dimostrazioni originali di Riemann").
Il Titchmarsh, si preoccupa di spiegare i passi principali di tali dimostrazioni, dando per acquisite conoscenze avanzate nella trattazione delle funzioni di analisi complessa e reale come, ad esempio, le serie e le trasformate di Fourier.
Sono certamente dei procedimenti *straordinari* che si servono di svariate tecniche matematiche che però convergono alla stessa conclusione: l'equazione funzionale della ζ.

13. GLI ZERI DELLA ζ E L'IPOTESI DI RIEMANN

Questa è la sezione centrale di tutta la tesi. Dopo aver introdotto la funzione ζ di Riemann, averne sviscerato le caratteristiche più importanti e descritto anche i due metodi utilizzati da Riemann stesso per dimostrare l'esistenza di un'equazione funzionale, ora passeremo agli zeri della ζ, per poi concludere la sezione con l'ipotesi di Riemann.

13.1 LA FUNZIONE ξ DI RIEMANN

Prima di passare all'analisi degli zeri della funzione ζ di Riemann, introduciamo un'altra funzione direttamente ricavabile dall'equazione funzionale della ζ stessa e della quale Riemann dà delle interessanti rappresentazioni.

13.1.1 La funzione ξ di Riemann

Nella scorsa sezione avevamo mostrato le due diverse dimostrazioni di Riemann riguardo all'equazione funzionale della ζ giungendo alla seguente conclusione (§12.3)

$$\zeta(s) = 2\Gamma(1-s)(2\pi)^{s-1}\sin\left(\frac{\pi s}{2}\right)\zeta(1-s), \qquad Re(s) < 0,$$

o, analogamente,

$$\zeta(1-s) = 2\Gamma(s)(2\pi)^{-s}\sin\left(\frac{\pi(1-s)}{2}\right)\zeta(s), \qquad Re(s) > 0.$$

Avevamo anche visto una certa simmetria in questa equazione, applicando alcune proprietà della funzione Γ:

$$\Gamma\left(\frac{s}{2}\right)\zeta(s)\pi^{-\frac{s}{2}} = \Gamma\left(\frac{1-s}{2}\right)\pi^{-\frac{1-s}{2}}\zeta(1-s).$$

Se analizziamo il primo membro di questa relazione, cioè

$$\Gamma\left(\frac{s}{2}\right)\zeta(s)\pi^{-\frac{s}{2}}$$

possiamo notare che questa è una meromorfa poiché è definita come prodotto di funzioni olomorfe le cui uniche singolarità sono di tipo polo. I poli (semplici) si trovano in corrispondenza di $s = 0$ (a causa della funzione Γ) e di $s = 1$ (a causa della ζ, lo dimostreremo nel dettaglio in seguito). Riemann, allora, moltiplica ambo i membri della relazione per $s(s-1)/2$, ottenendo ([9], §1.8)

$$\Gamma\left(\frac{s}{2}\right)\zeta(s)\pi^{-\frac{s}{2}}\frac{s(s-1)}{2} = \Gamma\left(\frac{1-s}{2}\right)\pi^{-\frac{1-s}{2}}\zeta(1-s)\frac{s(s-1)}{2}.$$

Inoltre ingloba il termine "$s/2$" all'interno della funzione Γ sfruttando la relazione (§8.2)

$$\Gamma\left(\frac{s}{2}+1\right) = \frac{s}{2}\Gamma\left(\frac{s}{2}\right),$$

e anche

$$\Gamma\left(1+\frac{1-s}{2}\right) = \Gamma\left(\frac{3-s}{2}\right) = \frac{1-s}{2}\Gamma\left(\frac{1-s}{2}\right),$$

nel secondo membro, ricordando che la Γ è l'estensione del fattoriale naturale. Si ottiene, dunque

$$\Gamma\left(\frac{s}{2}+1\right)\zeta(s)\pi^{-\frac{s}{2}}(s-1) = \Gamma\left(\frac{3-s}{2}\right)\pi^{-\frac{1-s}{2}}\zeta(1-s)(-s),$$

cioè

$$(s-1)\Gamma\left(\frac{s}{2}+1\right)\zeta(s)\pi^{-\frac{s}{2}} = (-s)\Gamma\left(\frac{3-s}{2}\right)\pi^{-\frac{1-s}{2}}\zeta(1-s).$$

In questo modo giungiamo ad una relazione dove le funzioni al primo e al secondo membro sono analitiche e senza singolarità di tipo polo. Inoltre possiamo notare che, se sostituiamo s con $1-s$ al primo membro otteniamo il secondo

$$(1-s-1)\Gamma\left(\frac{1-s}{2}+1\right)\zeta(1-s)\pi^{-\frac{1-s}{2}} = (-s)\Gamma\left(\frac{3-s}{2}\right)\zeta(1-s)\pi^{-\frac{1-s}{2}},$$

da cui deduciamo che resta nuovamente la simmetria rispetto a $Re(s) = 1/2$.
Riemann, dunque, definisce la funzione $\xi(s)$

$$\xi(s) = (s-1)\Gamma\left(\frac{s}{2}+1\right)\zeta(s)\pi^{-\frac{s}{2}},$$

la quale è una funzione intera (cioè analitica in tutto $\mathbb{C}$ (§3.2.5)) e, per costruzione, soddisfa l'equazione funzionale

$$\xi(s) = \xi(1-s),$$

poiché

$$\xi(1-s) = (-s)\Gamma\left(\frac{3-s}{2}\right)\zeta(1-s)\pi^{-\frac{1-s}{2}},$$

come visto in precedenza sostituendo s con $1-s$ per dimostrare tale simmetria rispetto alla retta $Re(s) = 1/2$.

13.1.2 Osservazioni importanti per la ξ

Prima di andare avanti, è bene analizzare alcune proprietà della ξ, soprattutto per quanto riguarda i suoi zeri e il legame tra questi e gli zeri della ζ.

(i) Innanzitutto ribadiamo che la funzione ξ così definita è una funzione intera.
Nella sua definizione, infatti, si è scelto di moltiplicare ambo i membri dell'equazione funzionale della ζ per $s(s-1)/2$ in modo da cancellarne i poli dovuti alla ζ (per $s = 1$) e alla Γ (per $s = 0$), mentre il dividere per 2 è una scelta estetica che evita coefficienti numerici.
Sapevamo che la funzione ζ ha un polo per $s = 1$, cosa che si poteva notare dalla formulazione tramite la formula della somma di Eulero (§11.2)

$$\zeta(s) = 1 + \frac{1}{s-1} - s\int_1^{\infty} \frac{\{t\}}{t^{s+1}}\,dt\,, \qquad Re(s) > 0.$$

Nei prossimi paragrafi, inoltre, vedremo che tale polo sarà un polo semplice, proprio a partire dalla formulazione dovuta ad Eulero.

(ii) L'equazione funzionale

$$\xi(s) = \xi(1-s),$$

mostra molto chiaramente la già decantata simmetria per $Re(s) = 1/2$. Tale simmetria vale anche per gli zeri della ξ: se $s_0 \in \mathbb{C}$ è una radice della ξ (cioè $\xi(s_0) = 0$), anche $1 - s_0$ è uno zero della funzione ξ.

(iii) Si può mostrare, inoltre, che la funzione ξ (e anche la ζ) sono simmetriche rispetto all'asse reale, cioè $\xi(\bar{s}) = \xi(s)$.

(iv) Passando, ora, agli zeri della ξ, la definizione

$$\xi(s) = (s-1)\Gamma\left(\frac{s}{2}+1\right)\zeta(s)\pi^{-\frac{s}{2}},$$

mostra chiaramente che la funzione ha gli stessi zeri della ζ per $Re(s) > 1/2$ in quanto l'unico zero di $(s-1)$ annulla il polo della $\zeta(s)$ mentre la funzione Γ e $\pi^{-\frac{s}{2}}$ non si annullano mai.

(v) Una questione più delicata riguarda $Re(s) < 1/2$. Tenendo fede all'equazione funzionale, si ha

$$\xi(1-s) = (-s)\Gamma\left(\frac{3-s}{2}\right)\zeta(1-s)\pi^{-\frac{1-s}{2}},$$

anche qui è chiaro che gli zeri della ξ sono collegati in qualche modo con quelli della ζ.

Questo legame, tuttavia, non è immediato come quello per $Re(s) > 1/2$ e necessita di ulteriore analisi.

Osserviamo che $\zeta(1-s)$ si annulla per $s = -(2k+1)$, per k intero positivo. Sono questi gli zeri banali della funzione ζ, di cui avevamo parlato in precedenza (§12.3.2), sottolineando $\zeta(-2n) = 0$, per n intero positivo.

Tuttavia $\xi(1-s)$ non si annulla per $s = -(2k+1)$, poiché gli zeri banali della ζ si "cancellano" con i poli della funzione Γ in prossimità degli stessi punti (§8.3).

Possiamo concludere che la funzione ξ ha gli stessi zeri della ζ, eccetto quelli banali: un fatto importante che è alla base stessa della ragionevolezza dell'ipotesi di Riemann. Se, infatti, si riesce a dimostrare che la funzione ξ ha infiniti zeri – risultato trovato da Hadamard nel 1893, come vedremo nei prossimi paragrafi – si conclude automaticamente che la funzione ζ ha infiniti zeri non banali.

(vi) Per quanto riguarda il calcolo dei valori della funzione ξ ci si serve dell'equazione funzionale della stessa e della sua semplicità, ricavando $\xi(1-s)$ a partire da $\xi(s)$. Alcuni valori della $\xi(s)$ sono ([28])

- $\xi(1) = 1/2$
- $\xi(2) = \frac{\pi}{6}$
- $\xi(4) = \frac{\pi^2}{15}$

dai quali, tramite l'equazione funzionale $\xi(s) = \xi(1-s)$ si deduce

- $\xi(0) = \frac{1}{2}$
- $\xi(-1) = \frac{\pi}{2}$

- $\xi(-3) = \frac{\pi^2}{15}$.

13.1.3 Motivazioni della ξ

Ci si può chiedere, a questo punto, perché passare dalla ζ alla ξ. Riemann è stato uno dei pionieri (dopo Gauss) nel campo delle equazioni di variabile complessa e i suoi risultati più apprezzati riguardano la loro rappresentazione (c'è tutta una geometria "Riemanniana" che aiuta a rappresentare le funzioni di una variabile complessa) e la loro trattazione mediante estensioni analitiche, trasformate di vario tipo e altro.
La differenza tra la ξ e la ζ è, soprattutto, una differenza di completezza e di perfezione quasi artistica. La ξ, infatti, completa la funzione ζ annullando quel "fastidioso" polo per $s = 1$ e la semplifica graficamente trasformando l'uguaglianza

$$\zeta(s) = 2\Gamma(1-s)(2\pi)^{s-1}\sin\left(\frac{\pi s}{2}\right)\zeta(1-s), \qquad Re(s) < 0,$$

che, a parte la simmetria, coinvolge pur sempre un prodotto di vari fattori, in un semplice

$$\xi(s) = \xi(1-s),$$

sicuramente più gradevole anche per la vista.
Troppo spesso, infatti, si guarda alla matematica come un insieme "incomprensibile" di simboli alfanumerici il cui unico scopo sembrerebbe quello di complicare una realtà fornendo dimostrazioni incomprensibili e problemi fuori da ogni logica concreta.
Senza elencare le motivazioni che contraddicono la conclusione precedente, anticipiamo che, nella sezione dedicata al Teorema dei Numeri Primi (§Appendice III), avremo modo di constatare che così non è, perché vedremo che, dopo la prima dimostrazione dello stesso, ne sono state date altre, più semplici, più "elementari" (cioè capaci di evitare tecniche particolarmente complicate di analisi complessa), a manifestare come nella matematica si cerca anche la semplificazione oltre che il mero risultato.

13.1.4 Rappresentazione di Riemann per la ξ

Dopo aver introdotto la funzione ξ tramite l'equazione funzionale della ζ, Riemann trova un'altra rappresentazione della funzione ξ che ora andremo ad analizzare più nel dettaglio. Ricordiamo che, nella seconda dimostrazione dell'equazione funzionale della ζ si era giunti al seguente risultato (§12.3.4):

$$\pi^{\frac{s}{2}}\Gamma\left(\frac{s}{2}\right)\zeta(s) = \int_1^\infty \left(x^{-\frac{s}{2}-\frac{1}{2}} + x^{\frac{s}{2}-1}\right)\psi(x)dx + \frac{1}{s(s-1)},$$

ovvero

$$\xi(s)\frac{2}{s(s-1)} = \int_1^\infty \left(x^{-\frac{s}{2}-\frac{1}{2}} + x^{\frac{s}{2}-1}\right)\psi(x)dx + \frac{1}{s(s-1)},$$

cioè

$$\xi(s) = \frac{1}{2} + \frac{s(s-1)}{2}\int_1^\infty \left(x^{-\frac{s}{2}-\frac{1}{2}} + x^{\frac{s}{2}-1}\right)\psi(x)dx.$$

A questo punto ricordiamo la formula di integrazione per parti

$$\int_a^b f(x)g'(x)dx = \int_a^b \frac{d}{dx}[f(x)g(x)]dx - \int_a^b f'(x)g(x)dx\,;$$

considerando

$$f(x) = x^{-\frac{s}{2}-\frac{1}{2}} + x^{\frac{s}{2}-1}, \qquad g(x) = \psi(x),$$

otteniamo, per $s \in \mathbb{C}$,

$$\begin{aligned}
\xi(s) &= \frac{1}{2} + \frac{s(s-1)}{2}\int_1^\infty \left(x^{-\frac{s}{2}-\frac{1}{2}} + x^{\frac{s}{2}-1}\right)\psi(x)dx \\
&= \frac{1}{2} - \frac{s(1-s)}{2}\int_1^\infty \left(x^{(1-s)/2} + x^{s/2}\right)\psi(x)\frac{dx}{x} \\
&= \frac{1}{2} - \frac{s(1-s)}{2}\int_1^\infty \frac{d}{dx}\left[\psi(x)\left(\frac{x^{(1-s)/2}}{(1-s)/2} + \frac{x^{s/2}}{s/2}\right)\right]dx \\
&+ \frac{s(1-s)}{2}\int_1^\infty \psi'(x)\left(\frac{x^{(1-s)/2}}{(1-s)/2} + \frac{x^{s/2}}{s/2}\right)dx \\
&= \frac{1}{2} - \frac{s(1-s)}{2}\psi(x)\left(\frac{x^{(1-s)/2}}{(1-s)/2} + \frac{x^{s/2}}{s/2}\right)\Bigg|_1^\infty \\
&+ \int_1^\infty \psi'(x)\left(sx^{(1-s)/2} + (1-s)x^{s/2}\right)dx \\
&= \frac{1}{2} + \frac{s(1-s)}{2}\psi(1)\left(\frac{2}{s} + \frac{2}{1-s}\right) + \int_1^\infty \psi'(x)\left(sx^{(1-s)/2} + (1-s)x^{s/2}\right)dx \\
&= \frac{1}{2} + \psi(1) + \int_1^\infty \psi'(x)\left(sx^{(1-s)/2} + (1-s)x^{s/2}\right)dx.
\end{aligned}$$

In questo calcolo si è utilizzato il teorema fondamentale del calcolo integrale

$$\int_a^b f'(x)dx = f(b) - f(a),$$

e, inoltre, il fatto che $\psi(x)$, definita come somma di esponenziali negativi con $x \in \mathbb{R}$ (§12.3.3), per $x \to +\infty$ (ricordiamo che $s \in \mathbb{C}$ è fissato) soddisfa:

$$\psi(x)\left(\frac{x^{(1-s)/2}}{(1-s)/2} + \frac{x^{s/2}}{s/2}\right) \xrightarrow{x\to+\infty} 0.$$

Andiamo avanti vedendo come Riemann riesce ad applicare ancora il teorema fondamentale del calcolo integrale e l'integrazione per parti per semplificare anche il secondo integrale.

$$\begin{aligned}\xi(s) &= \frac{1}{2} + \psi(1) + \int_1^\infty \psi'(x)\left(sx^{(1-s)/2} + (1-s)x^{s/2}\right)dx \\ &= \frac{1}{2} + \psi(1) + \int_1^\infty x^{3/2}\psi'(x)\left(sx^{-x/2-1} + (1-s)x^{(s-1)/2-1}\right)dx \\ &= \frac{1}{2} + \psi(1) + \int_1^\infty \frac{d}{dx}\left[x^{3/2}\psi'(x)\left(-2x^{-s/2} - 2x^{(s-1)/2}\right)\right]dx \\ &- \int_1^\infty \frac{d}{dx}\left[x^{3/2}\psi'(x)\right]\left(-2x^{-s/2} - 2x^{(s-1)/2}\right)dx \\ &= \frac{1}{2} + \psi(1) - \psi'(1)[-2-2] \\ &- \int_1^\infty \frac{d}{dx}\left[x^{3/2}\psi'(x)\right]\left(-2x^{-s/2} - 2x^{(s-1)/2}\right)dx \\ &= \frac{1}{2} + \psi(1) + 4\psi'(1) - \int_1^\infty \frac{d}{dx}\left[x^{3/2}\psi'(x)\right]\left(-2x^{-s/2} - 2x^{(s-1)/2}\right)dx.\end{aligned}$$

Ricordiamo, ora, l'equazione funzionale per la funzione ψ di Jacobi (§12.3.3)

$$2\psi(x) + 1 = \frac{1}{\sqrt{x}}\left(2\psi\left(\frac{1}{x}\right) + 1\right).$$

Derivando rispetto a x, otteniamo

$$2\psi'(x) = -\frac{1}{2}x^{-\frac{3}{2}}\left(2\psi\left(\frac{1}{x}\right) + 1\right) + x^{-\frac{1}{2}}\left(2\psi'\left(\frac{1}{x}\right)\cdot\left(-\frac{1}{x^2}\right)\right),$$

in essa, ponendo $x = 1$, si ha

$$2\psi'(1) = -\frac{1}{2}\left(1 + 2\psi(1)\right) + \left(-2\psi'(1)\right),$$

cioè

$$\frac{1}{2} + \psi(1) + 4\psi'(1) = 0,$$

che si può sostituire direttamente nel calcolo che stavamo operando:

$$\begin{aligned}\xi(s) &= \frac{1}{2} + \psi(1) + 4\psi'(1) - \int_1^\infty \frac{d}{dx}\left[x^{3/2}\psi'(x)\right]\left(-2x^{-s/2} - 2x^{(s-1)/2}\right)dx \\ &= \int_1^\infty \frac{d}{dx}\left[x^{3/2}\psi'(x)\right]\left(2x^{(s-1)/2} + 2x^{-s/2}\right)dx.\end{aligned}$$

A questo punto vogliamo esprimere il termine

$$2x^{-s/2} + 2x^{(s-1)/2},$$

in funzione del coseno iperbolico.

Ricordiamo la definizione di coseno iperbolico reale

$$\cosh(y) = \frac{1}{2}(e^y + e^{-y}),$$

definizione che si può estendere naturalmente nel campo complesso, così come le formule di addizione per lo stesso (a partire dalle proprietà dell'esponenziale complesso) che utilizzeremo in seguito ponendo $s = \frac{1}{2} + it$ dopo aver trovato la formula per la ξ.

Nell'integrando, isoliamo il termine $2x^{-1/4}$ per poi attuare la strategia appena enunciata

$$\xi(s) = \int_1^\infty \frac{d}{dx}\left[x^{3/2}\psi'(x)\right]2x^{-1/4}\left(x^{s/2-\frac{1}{4}} + x^{-s/2+\frac{1}{4}}\right)dx$$
$$= \int_1^\infty \frac{d}{dx}\left[x^{3/2}\psi'(x)\right]2x^{-1/4}\left(e^{\left(\frac{s}{2}-\frac{1}{4}\right)\log(x)} + e^{\left(-\frac{s}{2}+\frac{1}{4}\right)\log(x)}\right)dx$$
$$= \int_1^\infty \frac{d}{dx}\left[x^{3/2}\psi'(x)\right]2x^{-1/4}2\cosh\left[\frac{1}{2}\left(s-\frac{1}{2}\right)\log(x)\right]dx$$
$$= 4\int_1^\infty \frac{d}{dx}\left[x^{3/2}\psi'(x)\right]x^{-1/4}\cosh\left[\frac{1}{2}\left(s-\frac{1}{2}\right)log(x)\right]dx.$$

Stiamo per arrivare alla fine del calcolo. Riprendendo la definizione di coseno iperbolico

$$\cosh(y) = \frac{1}{2}(e^y + e^{-y}),$$

e gli sviluppi dell'esponenziale (complesso) in serie di potenze

$$e^y = \sum_{n=0}^{\infty}\frac{y^n}{n!}, \qquad e^{-y} = \sum_{n=0}^{\infty}(-1)^n\frac{y^n}{n!},$$

si ottiene

$$\cosh(y) = \frac{1}{2}(e^y + e^{-y}) = \sum_{n=0}^{\infty}\frac{y^{2n}}{(2n)!},$$

in quanto, nella somma tra i due esponenziali, i termini ad indice dispari si annullano tra loro mentre quelli ad indici pari si sommano. Otteniamo, dunque

$$\xi(s) = 4\int_1^\infty \frac{d}{dx}\left[x^{3/2}\psi'(x)\right]x^{-1/4}\sum_{n=0}^{\infty}\frac{\left[\frac{1}{2}\left(s-\frac{1}{2}\right)\log(x)\right]^{2n}}{(2n)!}dx,$$

la quale, isolando i termini dipendenti solo da s (che si possono estrarre dall'integrale), diventa

$$\xi(s) = \sum_{n=0}^{\infty}\left(s-\frac{1}{2}\right)^{2n}a_{2n},$$

in cui

$$a_{2n} = 4\int_1^\infty \frac{d}{dx}\left[x^{3/2}\psi'(x)\right]x^{-1/4}\frac{\left(\frac{1}{2}\log(x)\right)^{2n}}{(2n)!}dx,$$

rappresenta una sequenza di termini non dipendenti da s.

13.1.5 Osservazioni sulla rappresentazione della $\xi(s)$

A partire dalla formula appena trovata, si possono fare delle osservazioni e dei collegamenti all'articolo originale di Riemann presente nell'Appendice I di questa tesi (tradotto in italiano).

(i) Possiamo notare che la rappresentazione della $\xi(s)$ appena trovata, cioè

$$\xi(s) = \sum_{n=0}^{\infty}\left(s-\frac{1}{2}\right)^{2n}a_n,$$

con a_n opportuno (descritto in precedenza) è una funzione che mostra ancora una volta la simmetria della $\xi(s)$ rispetto alla linea $Re(s) = 1/2$, la stessa vista per l'equazione funzionale della ξ.
Inoltre questa stessa rappresentazione è la serie di potenze (quindi la serie di Taylor (2.3.1)) della ξ centrata in $s_0 = 1/2$. Un altro modo per rimarcare come la ξ sia analitica in $\mathbb{C}$ (§3.2.5, §3.3.4).

(ii) Riemann osserva poi che, se nell'equazione

$$\xi(s) = 4\int_1^\infty \frac{d}{dx}\left[x^{3/2}\psi'(x)\right]x^{-1/4}\cosh\left[\frac{1}{2}\left(s-\frac{1}{2}\right)\log(x)\right]dx,$$

poniamo $s = \frac{1}{2} + it$, otteniamo

$$\xi\left(\frac{1}{2}+it\right) = 4\int_1^\infty \frac{d}{dx}\left[x^{3/2}\psi'(x)\right]x^{-1/4}\cos\left(\frac{t}{2}\log(x)\right)dx,$$

utilizzando le formule di somma per il coseno iperbolico. Questa formula è la stessa trovata da Riemann nel suo articolo di ricerca (§Appendice I) anche se, poi, Riemann, come vedremo nel paragrafo dedicato all'ipotesi, continua a trattare con $t \in \mathbb{C}$, invece di $t \in \mathbb{R}$ come la logica potrebbe lasciar supporre.

(iii) In alcuni testi si utilizza la notazione

$$\Xi(t) = \xi\left(\frac{1}{2}+it\right), \qquad t \in \mathbb{R}.$$

Tuttavia, notazione a parte, il matematico tedesco è piuttosto oscuro in questi paragrafi. Inizialmente, a partire da questa formula dice che essa si può sviluppare "in una serie che converge molto rapidamente" ma non dà stime o ragguagli su questo fatto.
Allude senz'altro allo sviluppo in serie di $\xi(s)$ appena trovato ma è stato Hadamard, più di trent'anni dopo, a dimostrare questa "rapida convergenza" ottenendo un'interessante "formula prodotto" per la ξ che vedremo in seguito ma che dimostreremo nell'Appendice IV. Inoltre Riemann non specifica nemmeno se questo fatto gli interessi a qualcosa o gli serva per qualche calcolo futuro.

13.1.6 Formula prodotto per la ξ e infinità degli zeri

Vedremo che alla base della "ragionevolezza" dell'ipotesi di Riemann, sta l'affermazione o, meglio, la dimostrazione del fatto che la funzione ξ ha infiniti zeri o, analogamente, che la funzione ζ di Riemann ammette infiniti zeri non banali.
Tale affermazione non è semplice da provare e richiederebbe intere sezioni per essere approfondita nel modo che merita. C'è tutta una teoria delle funzioni intere che, tra le varie proprietà delle stesse, dimostra anche che queste – se soddisfano determinate condizioni – ammettono un numero infinito di zeri.
La funzione ξ è una di queste e, dunque, ammette infiniti zeri.
Tuttavia non dimostreremo questo fatto ma lo daremo per buono: nella prossima sezione, dedicata ai teoremi di von Mangoldt, mostreremo che il numero degli zeri – in genere chiamato "densità degli zeri" – della ξ con parte immaginaria compresa tra 0 e T, con T reale positivo, è

$$\frac{T}{2\pi}\log\left(\frac{T}{2\pi}\right)-\frac{T}{2\pi}+O(\log(T)).$$

Un tale risultato, infatti, per $T \to +\infty$ dimostrerà automaticamente che la ξ ammette infiniti zeri.

Quindi, per ora, prenderemo per buono il fatto che la funzione ξ ammette infiniti zeri, dunque che la funzione ζ ammette infiniti zeri non banali. Inoltre, in seguito vedremo che la parte reale degli zeri della ξ è compresa tra 0 e 1 (per questo, nella densità non si nomina la parte reale).

Richiamiamo, ora, alcuni passaggi tratti dall'articolo di Riemann (§Appendice I) con qualche piccola nota tra parentesi quadra.

> <<Questa funzione [cioè la ξ, *n.d.A.*] è finita per tutti i valori finiti di t, e permette essa stessa di essere sviluppata in potenze di tt ["tt" vale "t^2", *n.d.A.*] con una serie convergente molto rapidamente. Il numero di radici di $\xi(t)=0$, la cui parte reale è tra 0 e T è approssimativamente uguale a
>
> $$\frac{T}{2\pi}\log\frac{T}{2\pi}-\frac{T}{2\pi};$$
>
> poiché l'integrale $\int d\log\xi(t)$, fatto in senso positivo intorno alla regione consistente tra i valori di t la cui parte immaginaria si trova $\frac{1}{2}i$ e $-\frac{1}{2}i$ e la cui parte reale sta tra 0 e T, è (fino ad una frazione dell'ordine di grandezza della quantità $\frac{1}{T}$) uguale a $\left(T\log\frac{T}{2\pi}-T\right)i$. [...] Se si indica con α tutte le radici dell'equazione $\xi(\alpha)=0$, si può esprimere $\log\xi(t)$ come
>
> $$\sum\log\left(1-\frac{tt}{\alpha\alpha}\right)+\log\xi(0)\,;$$
>
> e, dal fatto che la densità delle radici della quantità t cresce con t solo come $\log\frac{t}{2\pi}$, segue che questa espressione converge e diventa per un t infinito solo infinita come $t\log t$; così differisce dal $\log\xi(t)$ per una funzione di tt, che per un t finito resta continua e finita e, quando è divisa da tt, diventa infinitamente piccola per t infinito. Questa differenza è, di conseguenza, una costante, il cui valore può essere determinato ponendo $t=0$.>>
>
> (tratto dall'articolo di Riemann)

Queste sono tra le righe più oscure di tutto l'articolo del matematico tedesco.

Si può notare che in esse compare anche la stima degli zeri della ξ di cui abbiamo appena parlato e che dimostreremo nella sezione dedicata ai teoremi di von Mangoldt. Tuttavia, Riemann parla di "zeri reali" e di "parte immaginaria" poiché la sua funzione ξ è leggermente differente da quella accettata e studiata in seguito, come vedremo nel paragrafo a fine sezione dedicato all'ipotesi di Riemann.

Si è discusso molto del perché Riemann pose molta attenzione sull'equazione funzionale (fornendone, addirittura, due dimostrazioni, sebbene molto stringate), per poi sintetizzare molto la questione relativa al numero degli zeri e tirare in ballo la formula prodotto in due righe senza nemmeno accennarne una dimostrazione o una motivazione per un utilizzo futuro della stessa.

Tuttavia, come già detto, nel 1893 il matematico J. Hadamard, dimostrò quanto accennato da Riemann, concludendo

$$\xi(s) = \xi(0) \prod_{\rho} \left(1 - \frac{s}{\rho}\right),$$

nella quale il prodotto è esteso a tutti i $\rho \in \mathbb{C}$ tali che $\xi(\rho) = 0$.
Inoltre, Riemann enuncia la formula nel modo seguente ([9], §1.16) (come si può riscontrare nel suo articolo):

$$\sum \log\left(1 - \frac{tt}{\alpha\alpha}\right) + \log \xi(0),$$

in essa "$\xi(0)$ <u>non</u> <u>è</u> $\xi(0)$". In precedenza Riemann aveva posto

$$s = \frac{1}{2} + it, \qquad t \in \mathbb{C},$$

come accennato in precedenza e come riprenderemo ampiamente nella sezione dedicata ai teoremi di von Mangoldt. In questo passaggio (e solo in esso), con

$$\xi(0),$$

Riemann intende

$$\xi(t), \qquad t = 0,$$

nel quale $\xi(t) \neq \xi(s)$ poiché, proprio come dice in precedenza

$$\xi\left(\frac{1}{2} + it\right) = \xi(s).$$

Si può provare che quello $\xi(0)$ equivale, dunque, a $\xi(1/2)$ ed è per questo che Riemann non lo esplicita poiché $\xi(1/2)$ non è un numero razionale (o comunque facilmente riconducibile a irrazionali comuni come radicali, π, e o altro). Solo in questo paragrafo, con la ξ, Riemann intende la *sua* ξ, cioè $\xi(t)$ ottenuta ponendo $s = \frac{1}{2} + it$ con t complesso.
Dimostreremo questa formula nell'Appendice IV.
Da ora in poi, indicheremo con ρ le radici della ξ. La posizione di queste radici nel piano complesso sarà l'oggetto del resto di questa sezione, ma, soprattutto, dell'ipotesi di Riemann.
Come già detto, per ora daremo per buono il fatto che la funzione ξ possiede infiniti zeri per poi vederne una dimostrazione nella sezione dedicata ai teoremi di von Mangoldt nella quale forniremo anche una stima degli stessi (la stessa indicata da Riemann nel suo articolo).

13.2 GLI ZERI DELLE FUNZIONI ξ E ζ DI RIEMANN

In questa sottosezione arriviamo al momento cruciale di tutto il discorso: cioè l'ipotesi di Riemann. Qui, infatti, inizieremo a parlare degli zeri della funzione ζ e della funzione ξ, oltre a discutere di alcune interessanti proprietà della ζ non viste in precedenza.

13.2.1 Il punto $s = 1$ e la linea $Re(s) = 1$

Iniziamo proprio con uno dei problemi insiti nel DNA della ζ fin dalla sua prima definizione, cioè la discontinuità nel punto $s = 1$. Avevamo, infatti, visto che

$$\zeta(1) = \sum_{n=1}^{\infty} \frac{1}{n} \to \infty,$$

proprio a causa della definizione stessa della ζ come estensione della serie armonica generalizzata (§11.1). Proviamo, preliminarmente, un risultato già accennato in precedenza.

Lemma
La funzione $\zeta(s)$ ha un polo semplice per $s = 1$ di residuo 1.
Dimostrazione
Ricordiamo la rappresentazione della $\zeta(s)$ per $Re(s) > 0$ ricavata dalla formula della somma di Eulero (§11.2)

$$\zeta(s) = \frac{1}{s-1} - s\int_1^{\infty} \frac{\{t\}}{t^{s+1}}\,dt + 1.$$

Moltiplichiamo ambo i membri per $(s-1)$ per poi calcolare il limite per $s \to 1$:

$$\lim_{s\to 1} \zeta(s)(s-1) = \lim_{s\to 1}\left(1 - s(s-1)\int_1^{\infty} \frac{\{t\}}{t^{s+1}}\,dt + (s-1)\right) = 1.$$

In essa, infatti, l'unico problema potrebbe essere l'integrale ma, questo, è un $O(t^{-s}) = O(t^{-1})$, quindi limitato. Moltiplicando, dunque, questa quantità limitata per $(s-1)$ che è infinitesimo, il risultato tende a zero.
Il limite che abbiamo calcolato è proprio quello nella definizione del residuo nel caso di una singolarità di tipo polo semplice (§3.5.1).

Vediamo, ora, di analizzare la situazione per $Re(s) = 1$, a parte il caso particolare già considerato di $s = 1$. Le cose vanno decisamente meglio rispetto al caso $s = 1$, in quanto possiamo osservare che $s = 1$ è l'unico polo.
Vedremo anche, in seguito, che $\zeta(s) \neq 0$, per $Re(s) = 1$.
Consideriamo, dunque, $s = 1 + it$ con $t \in \mathbb{R}$ andando a sostituirlo nella formulazione della ζ ottenuta con la formula di somma di Eulero (§11.2):

$$\zeta(s) = \zeta(1+it) = \frac{1}{1+it-1} - (1+it)\int_1^{\infty} \frac{\{t\}}{t^{(1+it)+1}}\,dt + 1$$
$$= \frac{1}{it} - \int_1^{\infty} \frac{\{t\}}{t^{2+it}}\,dt + it\int_1^{\infty} \frac{\{t\}}{t^{2+it}}\,dt.$$

Nel caso in cui $t \neq 0$ – quindi $Re(s) = 1$, con $s \neq 1$ – possiamo notare che non ci sono problemi di singolarità.

- Il termine $1/it$ è una funzione analitica (poiché $t \in \mathbb{R}$ con $t \neq 0$) e non presenta poli.
- L'integrando nel primo integrale è un $O(t^{-2})$, dunque tutto l'integrale è un $O(t^{-1})$: una quantità limitata.
- Nel secondo integrale, vale lo stesso ragionamento appena fatto.

Il termine it a prodotto con lo stesso non dà nessun problema: oltre ad essere una funzione analitica, abbiamo anche che $it \cdot O(t^{-1}) = O(1)$, cioè una costante.

13.2.2 Sugli zeri banali e non banali della $\zeta(s)$

Inizieremo mostrando un risultato classico e abbastanza intuitivo che ha, però, delle ripercussioni notevoli sia sulla ζ che sulla ξ.

Teorema
$\zeta(s) \neq 0$, per $Re(s) > 1$.
Dimostrazione
La dimostrazione segue abbastanza banalmente dalla formula del prodotto di Eulero per la funzione $\zeta(s)$ (§10.1.6):

$$\zeta(s) = \prod_{p \text{ primo}} \frac{1}{1 - 1/p^s} = \prod_{p \text{ primo}} \frac{p^s}{p^s - 1}, \qquad Re(s) > 1.$$

Da questa formula, soprattutto nella seconda rappresentazione, si deduce la tesi del teorema. Infatti, non esiste nessun $s \in \mathbb{C}$ per cui $p^s = 0$ (per ogni p primo), cioè non esiste nessun s che annulla uno qualsiasi dei numeratori del prodotto infinito.
Quanto detto vale perché il prodotto in questione converge per $Re(s) > 1$ per definizione dalla formula stessa del prodotto di Eulero.
Gli s per cui $p^s = 1$ sono tali che $Re(s) = 0$, quindi non contemplati nel caso di questo teorema.

Osservazioni
Il risultato appena trovato si può estendere al semipiano complesso $Re(s) < 0$ (ad eccezione degli zeri banali) grazie all'equazione funzionale della ζ (§12.3.1)

$$\zeta(s) = 2\Gamma(1-s)(2\pi)^{s-1} \sin\left(\frac{\pi s}{2}\right) \zeta(1-s),$$

che ci consente di calcolare $\zeta(s)$ per $Re(s) < 0$ a partire dai valori di s con $Re(s) > 1$. Si era anche visto che questa formula ha una sua simmetria per $Re(s) = 1/2$ (§12.3.2): tale simmetria si era, in seguito, sviluppata nella definizione della funzione ξ ad inizio sezione.
Focalizziamoci nell'equazione funzionale:

$$\zeta(s) = 2\Gamma(1-s)(2\pi)^{s-1} \sin\left(\frac{\pi s}{2}\right) \zeta(1-s)$$

analizzandone ad uno a uno i termini al secondo membro, nell'ipotesi $Re(s) < 0$.

- $\zeta(1-s)$ è la funzione ζ definita nel modo "classico", poiché $Re(1-s) > 1$: dal lemma appena visto sappiamo che, in quella regione, questo termine non si annulla mai.
- $\Gamma(1-s)$ è la funzione Γ che, anch'essa, non si annulla mai (§8.3). L'unico dettaglio degno di nota sono i poli per $s = -n$ con n intero non negativo: tuttavia non ci interessa poiché $Re(1-s) > 0$.
- $(2\pi)^{s-1}$ è un'esponenziale e, come tale, mai nullo.

Resta, dunque, il termine

$$\sin\left(\frac{\pi s}{2}\right),$$

il quale – sempre tenendo conto che $Re(s) < 0$ – si annulla per $s = -2n$, con n intero positivo dando origine agli zeri banali (§12.3.2).
Tutti gli altri zeri della ζ, che sono infiniti, sono detti zeri non banali e vedremo che si trovano in $0 < Re(s) < 1$, una zona del piano complesso detta, in tal senso, striscia critica.
Inoltre, dai risultati appena trovati, si deduce elementarmente che la ξ non ha zeri, né per $Re(s) > 1$, né per $Re(s) < 0$.
Basta ricordare l'osservazione fatta in precedenza (§13.1.2), sugli zeri della ξ e sul loro collegamento con quelli della ζ. La funzione ξ, infatti, ha gli stessi zeri della ζ eccetto quelli banali.

13.2.3 $\zeta(s)$ non si annulla per $Re(s) = 1$

Passiamo, ora, ad analizzare un altro risultato importante per la funzione ζ, cioè la dimostrazione della non esistenza degli zeri nella retta $Re(s) = 1$.
L'intera dimostrazione si basa sulla seguente osservazione:

$$3 + 4\cos(\theta) + \cos(2\theta) = 3 + 4\cos(\theta) + (2\cos^2(\theta) - 1) = 2(1 + 2\cos(\theta) + \cos^2(\theta))$$
$$= 2(1 + \cos(\theta))^2 \geq 0, \qquad \forall\theta \in \mathbb{R}.$$

Possiamo, quindi, passare alla dimostrazione generale del teorema.

Teorema ([5], §8)
$\zeta(1 + it) \neq 0, \forall t \in \mathbb{R}$.
Dimostrazione
Partiamo dalla formula del prodotto di Eulero per arrivare alla seguente formulazione valida per $Re(s) > 1$

$$\log(\zeta(s)) = \log\left(\prod_{p \text{ primo}} \frac{1}{1 - p^{-s}}\right) = -\sum_{p \text{ primo}} \log(1 - p^{-s}) = \sum_{p \text{ primo}} \sum_{r=1}^{\infty} \frac{1}{rp^{rs}},$$

che nell'ultimo passaggio utilizza lo sviluppo di Taylor di $\log(1 - x)$, deducibile da quello di $\log(1 + x)$ (§1.3.2) semplicemente considerando $-x$ in luogo di x.
Possiamo, dunque, scrivere

$$\zeta(s) = \exp\left(\sum_{p \text{ primo}} \sum_{r=1}^{\infty} \frac{1}{rp^{rs}}\right),$$

nella quale, per semplicità di scrittura, abbiamo scritto $\exp(f(s))$ in luogo di $e^{f(s)}$. A questo punto possiamo passare al modulo, ricordando che $p^s = e^{s\log(p)} = e^{(\sigma+it)\log(p)}$, per $s = \sigma + it$:

$$|\zeta(\sigma+it)| = \exp\left(\sum_{p \text{ primo}} \sum_{r=1}^{\infty} \frac{\cos(rt\log(p))}{rp^{r\sigma}}\right).$$

Sfruttando questa relazione, vogliamo trovarne una analoga per

$$\zeta^3(\sigma)|\zeta(\sigma+it)|^4|\zeta(\sigma+2it)|.$$

Partiamo dai singoli termini, otteniamo

$$\zeta^3(\sigma) = |\zeta^3(\sigma)| = \exp\left(\sum_{p \text{ primo}} \sum_{r=1}^{\infty} \frac{3\cos(r\cdot 0\cdot\log(p))}{rp^{r\sigma}}\right) = \exp\left(\sum_{p \text{ primo}} \sum_{r=1}^{\infty} \frac{3}{rp^{r\sigma}}\right),$$

per il primo,

$$|\zeta(\sigma+it)|^4 = \exp\left(\sum_{p \text{ primo}} \sum_{r=1}^{\infty} \frac{4\cos(rt\log(p))}{rp^{r\sigma}}\right),$$

per il secondo e, per l'ultimo

$$|\zeta(\sigma+2it)| = \exp\left(\sum_{p \text{ primo}} \sum_{r=1}^{\infty} \frac{\cos(2rt\log(p))}{rp^{r\sigma}}\right).$$

A questo punto dobbiamo moltiplicarli insieme, ricordando la basilare proprietà dell'esponenziale, valida anche in campo complesso (§3.2.7)

$$e^z \cdot e^w = e^{z+w}, \qquad z, w \in \mathbb{C},$$

anche se nella difficoltà sta nel fatto che la sommatoria è infinita. Ci serviremo anche della seguente proprietà delle sommatorie (§1.2.2)

$$\sum_{i=n}^{m} a_i \pm \sum_{i=n}^{m} b_i = \sum_{i=n}^{m} (a_i \pm b_i),$$

valida, come già detto, anche nel caso di indici infiniti (a patto che siano gli stessi indici).

$$\begin{aligned}
&\zeta^3(\sigma)|\zeta(\sigma+it)|^4|\zeta(\sigma+2it)| \\
&\quad = \exp\left(\sum_{p \text{ primo}} \sum_{r=1}^{\infty} \frac{3}{rp^{r\sigma}}\right) \cdot \exp\left(\sum_{p \text{ primo}} \sum_{r=1}^{\infty} \frac{4\cos(rt\log(p))}{rp^{r\sigma}}\right) \\
&\quad \cdot \exp\left(\sum_{p \text{ primo}} \sum_{r=1}^{\infty} \frac{\cos(2rt\log(p))}{rp^{r\sigma}}\right) \\
&\quad = \exp\left(\sum_{p \text{ primo}} \sum_{r=1}^{\infty} \frac{3}{rp^{r\sigma}} + \sum_{p \text{ primo}} \sum_{r=1}^{\infty} \frac{4\cos(rt\log(p))}{rp^{r\sigma}}\right. \\
&\quad \left. + \sum_{p \text{ primo}} \sum_{r=1}^{\infty} \frac{\cos(2rt\log(p))}{rp^{r\sigma}}\right) \\
&\quad = \exp\left(\sum_{p \text{ primo}} \sum_{r=1}^{\infty} \frac{3 + 4\cos(rt\,log(p)) + \cos(2rt\log(p))}{rp^{r\sigma}}\right).
\end{aligned}$$

Dall'osservazione antecedente il teorema, deduciamo che l'esponente è sempre non negativo, concludendo

$$\exp\left(\sum_{p \text{ primo}} \sum_{r=1}^{\infty} \frac{3 + 4\cos(rt\, log(p)) + \cos(2rt \log(p))}{rp^{r\sigma}}\right) \geq \exp(0) = e^0 = 1,$$

dunque

$$\zeta^3(\sigma)|\zeta(\sigma + it)|^4|\zeta(\sigma + 2it)| \geq 1,$$

moltiplicando e dividendo il primo membro per $(\sigma - 1)^4$. Tenendo conto che $(\sigma - 1)^4 = |\sigma - 1|^4$ poiché $\sigma = Re(s) \in \mathbb{R}$, otteniamo

$$(\sigma - 1)^4\zeta^3(\sigma)\left|\frac{\zeta(\sigma + it)}{\sigma - 1}\right|^4 |\zeta(\sigma + 2it)| \geq 1,$$

da cui, dividendo ambo i membri per $\sigma - 1$, otteniamo

$$\big((\sigma - 1)\zeta(\sigma)\big)^3\left|\frac{\zeta(\sigma + it)}{\sigma - 1}\right|^4 |\zeta(\sigma + 2it)| \geq \frac{1}{\sigma - 1}.$$

Fissiamo, ora, t e facciamo tendere $\sigma \to 1$:

- $\big((\sigma - 1)\zeta(\sigma)\big)^3 \to 1$, come abbiamo visto nel lemma in (§13.2.1);
- $|\zeta(\sigma + 2it)| \to |\zeta(1 + 2it)|$.

Il nucleo della questione sta proprio nel secondo termine. Se supponiamo $\zeta(1 + it) = 0$, abbiamo

$$\left|\frac{\zeta(\sigma + it)}{\sigma - 1}\right|^4 = \left|\frac{\zeta(\sigma + it) - \zeta(1 + it)}{\sigma - 1}\right|^4 \xrightarrow{\sigma \to 1} |\zeta'(1 + it)|^4,$$

quest'ultimo per definizione usuale di derivata come limite del rapporto incrementale. Dunque, se $\sigma \to 1$, abbiamo

$$\big((\sigma - 1)\zeta(\sigma)\big)^3\left|\frac{\zeta(\sigma + it)}{\sigma - 1}\right|^4 |\zeta(\sigma + 2it)| \to |\zeta'(1 + it)|^4|\zeta(1 + 2it)|,$$

che è un valore finito, poiché la $\zeta(s)$ è analitica a parte $s = 1$ (quindi la sua derivata esiste ed è finita): al massimo sarà un valore nullo se anche $\zeta(1 + 2it) = 0$, ma non cambia il risultato. Tuttavia, al secondo membro

$$\frac{1}{\sigma - 1} \xrightarrow{\sigma \to 1} \infty,$$

che è in contraddizione con il risultato appena trovato per quanto riguarda il primo membro.
Si conclude che non esiste t tale per cui $\zeta(1 + it) = 0$, cioè che la $\zeta(s)$ non ha zeri per $Re(s) = 1$.

Corollario
La funzione ξ non ha zeri per $Re(s) = 1$.
Dimostrazione
Segue in maniera elementare dalla definizione stessa della ξ (§13.1.1):

$$\xi(s) = (s - 1)\Gamma\left(\frac{s}{2} + 1\right)\zeta(s)\pi^{-\frac{s}{2}}.$$

Osservando che:

- $(s - 1) \neq 0$ per $Re(s) = 1$, tranne il già citato $s = 1$ nel quale, però, lo zero annulla il polo (semplice) della $\zeta(s)$.
- $\Gamma\left(\frac{s}{2} + 1\right) \neq 0$, in quanto la Γ non ha zeri in generale.
- $\zeta(s) \neq 0$, per $Re(s) = 1$, per il teorema appena visto.

- $\pi^{-s/2} \neq 0$ per definizione stessa di esponenziale.

Corollario
La funzione ζ non ha zeri per $Re(s) = 0$.
Dimostrazione
Segue elementarmente dall'equazione funzionale della ζ (§12.3.2):

$$\zeta(1-s) = 2\Gamma(s)(2\pi)^{-s} \sin\left(\frac{\pi(1-s)}{2}\right)\zeta(s), \qquad Re(s) > 0.$$

Basta porre $s = 1 + it$ in essa e fare le stesse considerazioni utilizzate per dimostrare il precedente corollario.

Corollario
La funzione ξ non ha zeri per $Re(s) = 0$.
Dimostrazione
Si ricava immediatamente dall'equazione funzionale della ξ (§13.1.2)

$$\xi(s) = \xi(1-s),$$

e dal fatto che $\xi(s) = 0$, per $Re(s) = 1$, ricordando che se $Re(s) = 1$, $Re(1-s) = 0$.

13.3 L'IPOTESI DI RIEMANN

Siamo arrivati al punto centrale di tutta la tesi, cioè l'ipotesi di Riemann; prima di enunciarla, però, è opportuno fare una panoramica della situazione.

13.3.1 Gli zeri della ζ e quelli della ξ: striscia critica

Ricapitoliamo la situazione in base ai teoremi appena visti; alla maniera di Riemann, è più semplice farlo per la ξ per poi tornare, a ritroso, verso la funzione ζ.
Diamo, quindi, un ordine a tutto il discorso.

1. $\xi(s)$ ha infiniti zeri. Lo abbiamo preso per buono parlando della formula prodotto per la stessa dimostrata da Hadamard (§13.1.6) che dimostreremo nell'Appendice IV. Nella sezione dedicata ai teoremi di von Mangoldt daremo una stima di questi zeri dimostrando, dunque, anche che sono infiniti. Da questa considerazione seguiva automaticamente che $\zeta(s)$ ha infiniti zeri non banali (§13.1.2).
2. $\xi(s)$ non ha zeri né per $Re(s) > 1$, né per $Re(s) < 0$ (§13.2).
3. $\xi(s)$ non ha zeri né per $Re(s) = 1$, né per $Re(s) = 0$ per gli ultimi due corollari considerati.

Si può concludere che gli zeri della ξ, dunque, si trovano nella zona $0 < Re(s) < 1$ che, per questo, prende il nome di "striscia critica" (in inglese "critical strip") Figura 13.1.

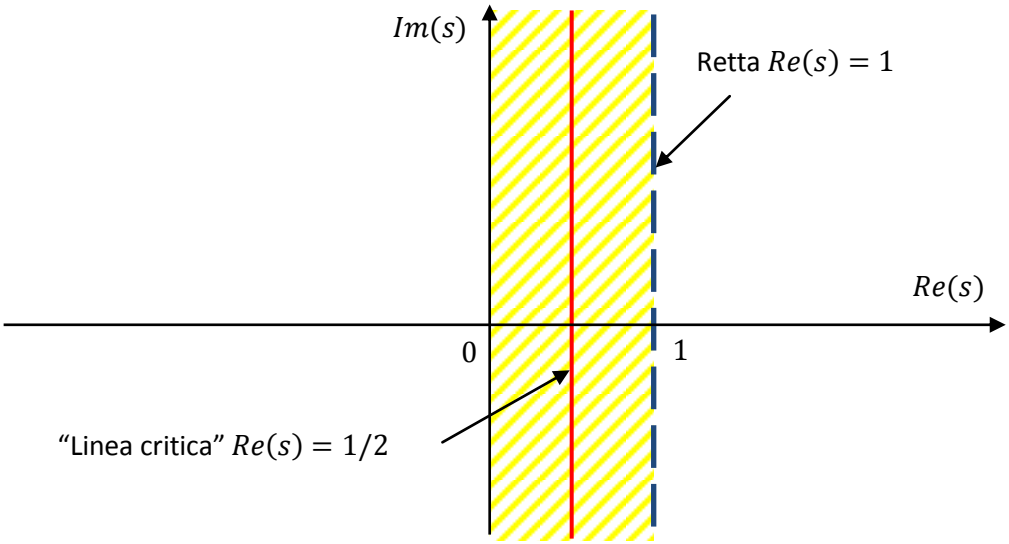

Figura 13.1. Striscia critica (tratteggiata in giallo) e linea critica (in rosso) nel piano complesso.

Si può vedere, in Figura 13.1, la rappresentazione della striscia critica $0 < Re(s) < 1$ e della linea critica $Re(s) = 1/2$ della quale parleremo a breve.

13.3.2 Dall'articolo di Riemann all'ipotesi

Riprendiamo un altro passo dell'articolo di Riemann (§Appendice I).

<< Io ora pongo $s = \frac{1}{2} + ti$ […]

$$\xi(t) = 4\int_1^{\infty} \frac{d\left(x^{\frac{3}{2}}\psi'(x)\right)}{dx} x^{-\frac{1}{4}} \cos\left(\frac{1}{2} t \log x\right) dx.$$

Questa funzione è finita per tutti i valori finiti di t, e permette essa stessa di essere sviluppata in potenze di tt ["tt" vale "t^2", *n.d.A.*] con una serie convergente molto rapidamente. Dal fatto che, per un valore di s la cui parte reale è più grande di 1, $\log \zeta(s) = -\sum \log(1 - p^{-s})$ resta finito, e dal fatto che la stessa cosa vale per i logaritmi degli altri fattori di $\xi(t)$, segue che la funzione $\xi(t)$ può solo annullarsi se la parte immaginaria di t si trova tra $\frac{1}{2}i$ e $-\frac{1}{2}i$.

[…] … è molto probabile che tutte le radici [non banali, sta parlando della ξ, *n.d.A.*] sono reali. Certamente ci si augura una piccola dimostrazione qui; nel frattempo io ho temporaneamente messo da parte la ricerca per questo dopo qualche futile tentativo di sfuggita, così come sembra non necessario per il prossimo obiettivo della mia indagine.>>

(tratto dall'articolo di Riemann)

La funzione esaminata da Riemann, cioè $\xi(t)$, non è da confondere con la già citata $\Xi(t)$, nella quale t è reale. Riemann, infatti, continua a considerare t complesso e conclude che tutti gli zeri sono per

$$-\frac{1}{2} < Im(t) < \frac{1}{2}.$$

Se, infatti, ponessimo $t = x + iy$, Riemann dice

$$-\frac{1}{2} < y < \frac{1}{2},$$

dunque

$$s = \frac{1}{2} + ti = \frac{1}{2} + i(x + iy) = \frac{1}{2} + y + ix,$$

in cui, tenendo conto della relazione per la y, si conclude $0 < Re(s) < 1$ che è proprio quanto trovato nel precedente paragrafo.

Tuttavia, Riemann va oltre e conclude che "è molto probabile che tutte le radici sono reali", cioè $Im(t) = 0$ nel suo articolo. Se, andassimo a sostituire $Im(t) = 0$ con $y = 0$, otterremo

$$s = \frac{1}{2} + ti = \frac{1}{2} + i(x + iy) = \frac{1}{2} + ix,$$

dunque $Re(s) = 1/2$, retta che prende il nome di "linea critica" (Figura 13.1).

L'affermazione di Riemann, infatti, equivale a dire che è probabile che tutte le radici della funzione ξ si trovino lungo la linea critica $Re(s) = 1/2$.

Ora, dalla funzione ξ, si può risalire facilmente, con analoghe conclusioni, proprio alla ζ per la definizione stessa di ξ

$$\xi(s) = (s-1)\Gamma\left(\frac{s}{2} + 1\right)\zeta(s)\pi^{-\frac{s}{2}},$$

dal momento che, nella striscia critica, le funzioni ξ e ζ hanno gli stessi zeri (in quanto, come già detto, tutti gli altri termini coinvolti in quel prodotto non si annullano per $0 < Re(s) < 1$).

La conclusione è dunque analoga e ci consente di enunciare la seguente

Ipotesi di Riemann

Tutte gli zeri non banali della funzione $\zeta(s)$ sono tali che $Re(s) = 1/2$.

13.3.3 Osservazioni/Conclusioni sull'ipotesi di Riemann

Si possono fare osservazioni più o meno filosofiche o matematiche riguardo ad uno dei problemi che ha tenuto impegnate le maggiori menti matematiche dell'ultimo secolo (e mezzo), cioè l'ipotesi (o congettura) di Riemann.

(i) Per quanto ora possa sembrare banale, così enunciata, in realtà per arrivare ad essa siamo dovuti passare attraverso l'analisi matematica (I, II e complessa), oltre che TDN e TADN.

(ii) E' lecito pensare che Riemann, in questa conclusione, sia stato influenzato proprio dalla simmetria della ξ lungo l'asse $Re(s) = 1/2$ che, come abbiamo visto, implica che, qualora $s_0 \in \mathbb{C}$ sia tale che $\xi(s_0) = 0$, anche $1 - s_0$ è tale che $\xi(s_0) = 0$.

(iii) L'ipotesi di Riemann riguarda la ξ, anche se poi si estende, come detto, in maniera naturale alla ζ ed oggi è quella espressa nel paragrafo precedente la formulazione accettata per la stessa.

(iv) Vedremo (§Appendice II) che Hardy dimostrò nel 1914 che la funzione $\xi(s)$ possiede infiniti zeri lungo la linea critica $Re(s) = 1/2$ (quindi anche per la ζ vale altrettanto).
Tuttavia la questione non è risolta perché dire "esistono infiniti zeri per $Re(s) = 1/2$", non esclude la presenza di altri zeri in altre regioni della striscia critica.

(v) L'ipotesi di Riemann non è ancora dimostrata.
Era stata inclusa, inizialmente, nella famosa lista dei 23 problemi irrisolti della matematica da parte di Hilbert nel 1900 ([21]). Tuttavia, al contrario di molti degli altri problemi, tale questione rimase irrisolta per tutto il secolo successivo e anzi venne inserita nuovamente nella lista dei così detti 7 problemi del millennio (le principali questioni matematiche all'alba del 2000) ([22]).

Nelle prossime sezioni ne vedremo altri sviluppi, trattando teoremi importanti come la formula di Riemann-von Mangoldt o l'equazione funzionale approssimata, per poi concludere la tesi con le conseguenze dell'ipotesi di Riemann in svariati campi della TDN, tra cui l'esistenza di una formula per la funzione enumerativa dei primi $\pi(x)$.
Tuttavia, prima di concludere, daremo uno sguardo alla locazione dei primi zeri non banali – in ordine di modulo – della funzione ζ (e dunque della ξ) per poi concludere con due rappresentazioni grafiche ricorrenti della stessa.

13.3.4 I primi zeri non banali della funzione ζ

In rete si possono trovare svariate tavole riguardo alla locazione dei primi zeri (non banali) della funzione ζ. Dall'inizio del secolo scorso – anno in cui si è iniziato a conteggiare metodicamente tali zeri – fino ad oggi sono stati calcolati tutti gli zeri con valori della parte immaginaria di ordine fino a 10^{13} ([27]).
Tuttavia sono presenti altri casi – isolati – di calcolo in regioni ancora più distanti dall'origine, la maggior parte dei quali sono dovuti al lavoro (anche teorico) del matematico Odlyzko ([31]) che, in collaborazione con Schönhage, ha saputo creare un algoritmo piuttosto efficiente per il calcolo dei valori della zeta (si parla di $O(T^{1+\varepsilon})$ per un input con valore T della parte immaginaria) ([8]).
Questo stesso algoritmo è alla base dei moderni calcoli, gli stessi che hanno permesso di giungere al quadro completo degli zeri (non banali) della ζ per valori con coefficiente immaginario da 0 a 10^{13} (quindi anche da 0 a -10^{13}, poiché $\zeta(\bar{s}) = \zeta(s), \forall s \in \mathbb{C}$).
Da questi calcoli, si evince che gli zeri non banali della funzione ζ sono

$$\rho \in \mathbb{C}, \qquad \rho = \frac{1}{2} + it, \qquad t \in \mathbb{R},$$

cioè sembrano tutti confermare l'ipotesi di Riemann.

Tuttavia, un proverbio piuttosto comune in matematica dice che "N indizi non formano una prova" (per quanto possa essere grande N) e si continua a lavorare in questo campo, sia per ulteriori conferme, sia per eventuali smentite.

- Se l'ipotesi di Riemann è vera, questi calcoli non bastano per provarla: forniscono tavole sempre più ampie degli zeri, ma non una dimostrazione.
- Se l'ipotesi di Riemann è falsa, ammesso che qualcuno riesca a dimostrarne la falsità, è ragionevole supporre che i calcoli, prima o poi, finiranno per individuare uno zero $\tilde{\rho}$ (non banale) della funzione ζ tale che $Re(\tilde{\rho}) \neq 1/2$.

Vediamo, dunque, l'ubicazione dei primi zeri non banali della funzione ζ (quindi dei primi zeri della funzione ξ). Come già detto, $\zeta(s) = \zeta(\bar{s})$, quindi una tale tabella riassuntiva comprende automaticamente i valori

$$\bar{\rho} = \frac{1}{2} - it,$$

a partire dai valori $\rho = 1/2 + it$. In rete si trovano molte tabelle degli zeri e, spesso, in esse, è indicato solamente il valore della parte immaginaria $t \in \mathbb{R}$ degli zeri, in quanto, fino ad ora, essi hanno tutti parte reale $1/2$.

Numero dello zero (in ordine di distanza dall'origine).	Valore di t (troncato alle prime 10 cifre decimali
1	14,1347251417
2	21,0220396387
3	25,0108575801
4	30,4248761258
5	32,9350615877
6	37,5861781588
7	40,9187190121
8	43,3270732809
9	48,0051508811
10	49,7738324776

Questa tabella è adattata a partire dall'analoga di Odlyzko (tuttavia con precisione fino alla millesima cifra decimale) in ([31]).

Come detto, tale tabella va intesa nel modo seguente: l'$n-$esimo zero (non banale) della funzione ζ è $1/2 + it$, nel quale t è il punto preso dalla tabella stessa. Per esempio, il sesto zero è (approssimativamente) il numero complesso

$$\frac{1}{2} + 37{,}5861781588i,$$

cioè

$$\zeta\left(\frac{1}{2} + 37{,}5861781588\ldots i\right) = 0.$$

13.3.5 Rappresentazioni grafiche della ζ

Concludiamo questa sezione analizzando due comuni rappresentazioni grafiche della funzione ζ: riporteremo, infatti, due tra i grafici più ricorrenti nei quali cercheremo di individuare le caratteristiche principali della ζ tra cui l'ubicazione degli zeri non banali.
Entrambe le immagini sono prese da wikipedia ([15]).

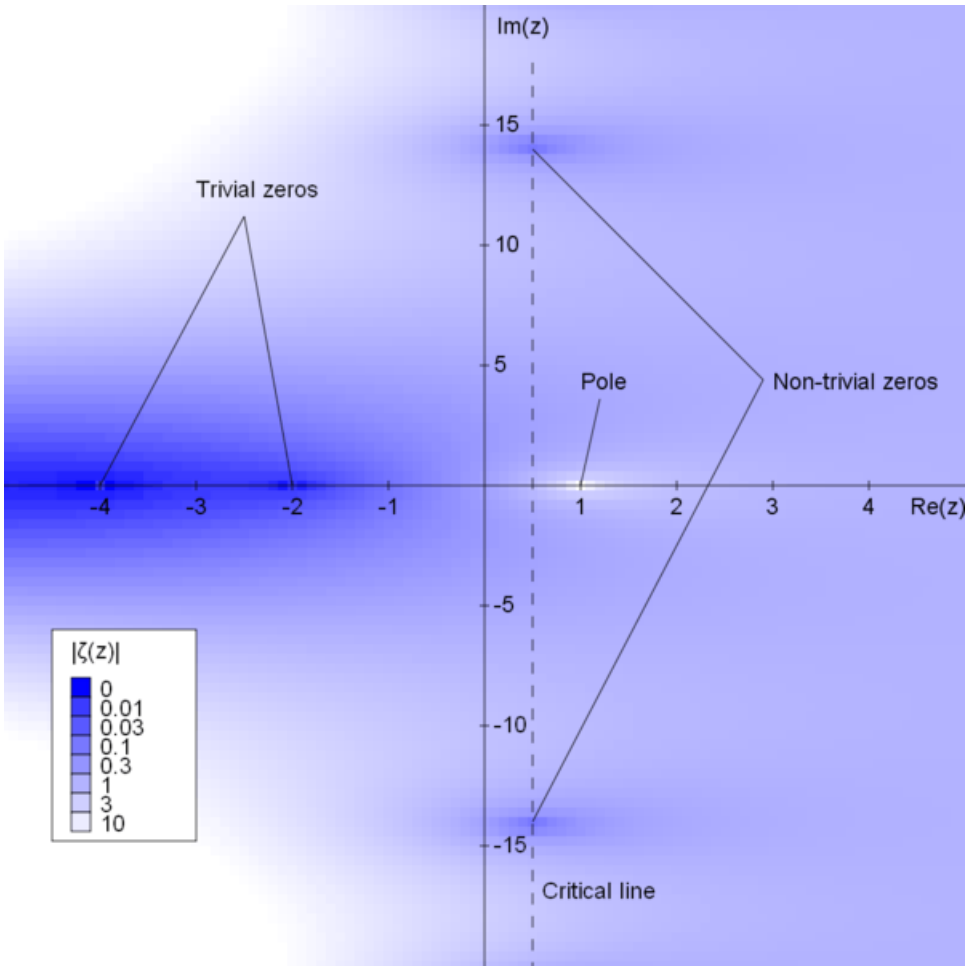

Figura 13.2. Modulo della $\zeta(z)$ per $-5 \leq Re(z) \leq 5$, $-5 \leq Im(z) \leq 5$.

La Figura 13.2 è un grafico bidimensionale a variazione di colori, di cui si è parlato nella sezione dedicata, appunto, ai grafici di funzione (§4).
La legenda ci mostra il valore di $|\zeta(z)|$ rappresentato da una gradazione differente di blu (dallo scuro per quelli di modulo minore al chiaro per quelli di modulo maggiore). Possiamo notare che ci sono 2 zeri non banali ("non-trivial zeros") – tra l'altro simmetrici rispetto all'origine poiché $\zeta(z) = \zeta(\bar{z})$ – e i primi due zeri banali ("trivial-zeros") in prossimità di valori interi negativi pari.
Si può vedere anche la linea critica ("critical line", $Re(z) = 1/2$) e il polo in prossimità di $z = 1$.
La prossima immagine, invece, ha dei fini più artistici che concreti: tra l'altro è quella utilizzata da Derbyshire come copertina oltre che come indicatore dei vari capitoli ([7]).

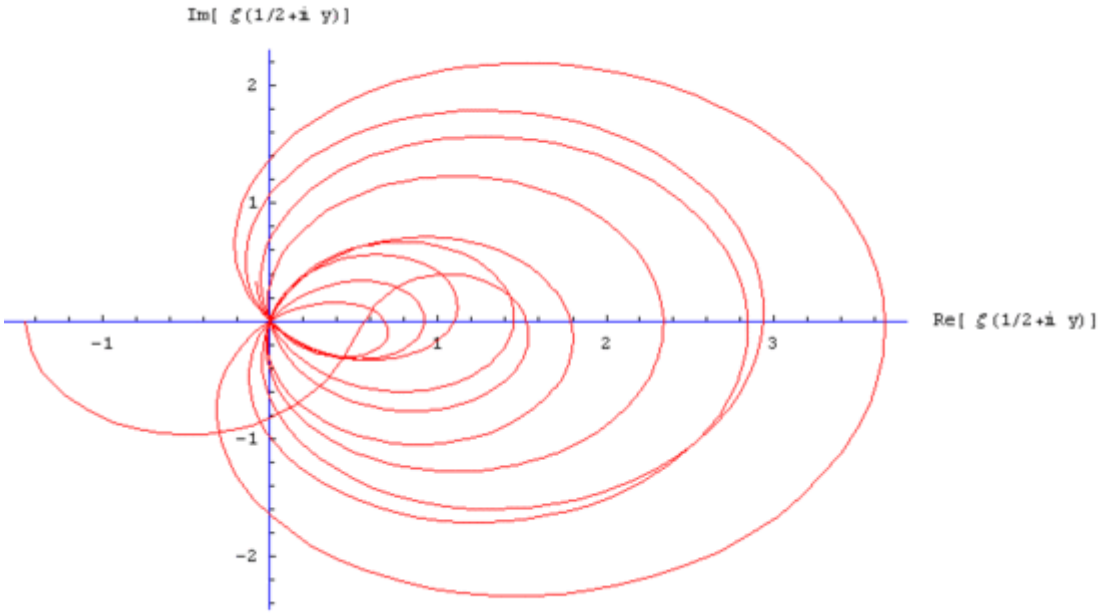

Figura 13.3. $Re\big(\xi(1/2+it)\big)$, *per* $0 \leq t \leq 50$.

Questa immagine mostra $Re\big(\xi(1/2+it)\big)$, con $t = Im(s)$ compreso tra 0 e 50. Il valore corrispondente a $t = 0$ è quello negativo a sinistra (poiché $\zeta(1/2) \cong -1{,}460$ ([29])), in seguito, per valori crescenti di $Im(s)$, $Re\big(\xi(1/2+it)\big)$ si sposta lungo la linea rossa (in Figura 13.3), creando quest'immagine piuttosto artistica.

In rete, inoltre, è possibile trovare molti grafici tridimensionali relativi al modulo della ζ in varie regioni del piano complesso (soprattutto nei pressi della striscia critica). Ce ne sono vari, da diverse angolazioni, e ci si può rapportare – come analisi degli stessi – a quanto detto riguardo ai grafici tridimensionali nell'apposita sezione (§4).

14. TEOREMI DI VON MANGOLDT (STIMA DEGLI ZERI E FORMULA ESPLICITA)

Questa sezione sarà dedicata a due importanti risultati ottenuti dal matematico von Mangoldt a partire dal lavoro di Riemann.
Il primo di essi riguarda la distribuzione degli zeri lungo la striscia critica accennato da Riemann e fu dimostrato da von Mangoldt circa mezzo secolo dopo.
In seguito proveremo la formula di Perron che sarà un tassello fondamentale per concludere con ultimo importante risultato, cioè la formula esplicita per la funzione ψ di Chebyshev a partire dagli zeri della ξ (cioè gli zeri non banali della ζ).

14.1 TEOREMA DI RIEMANN-VON MANGOLDT

14.1.1 Introduzione

In questa sottosezione dimostreremo la stima per gli zeri della funzione ξ. Nel suo articolo (§Appendice I), Riemann, parla di tale stima, mostrando anche alcuni passi della dimostrazione

> << Il numero di radici di $\xi(t) = 0$, la cui parte reale è tra 0 e T è approssimativamente uguale a
> $$\frac{T}{2\pi}\log\frac{T}{2\pi} - \frac{T}{2\pi};$$
> poiché l'integrale $\int d\log\xi(t)$, fatto in senso positivo intorno alla regione consistente tra i valori di t la cui parte immaginaria si trova $\frac{1}{2}i$ e $-\frac{1}{2}i$ e la cui parte reale sta tra 0 e T, è (fino ad una frazione dell'ordine di grandezza della quantità $\frac{1}{T}$) uguale a $\left(T\log\frac{T}{2\pi} - T\right)i$; questo integrale comunque è uguale al numero delle radici di $\xi(t) = 0$ che giacciono in questa regione, moltiplicati per $2\pi i$.>>
>
> (Tratto dall'articolo di Riemann)

Tuttavia, come abbiamo già detto nella sezione dedicata all'ipotesi (§13.1.6), il matematico tedesco trattava una versione "differente" della ξ rispetto a quella usuale proprio perché partiva dal presupposto di considerare

$$\xi\left(\frac{1}{2}+it\right), \qquad t \in \mathbb{C},$$

nel quale $t = x + iy$ è visto come un numero complesso qualsiasi e una tale scrittura aveva il pregio di dare più eleganza all'ipotesi di Riemann in sé e alla locazione degli zeri.
Questa è una scrittura che contrasta con ciò che siamo abituati a vedere poiché, generalmente, indicando

$$\xi\left(\frac{1}{2}+it\right),$$

si pensa a $\xi(s)$ nel quale si fissa la parte reale a $1/2$. Riemann, invece, considera t come variabile complessa in modo da avere, nella sua ipotesi, radici reali (§13.3.2).
Tornando all'eleganza, la ξ stessa è un esempio di una tale necessità artistica: essa nasce come necessità di "completare" in senso analitico la ζ che di per sé ha due "fastidiose" singolarità in $z = 1$ e $z = 0$ (per l'equazione funzionale). La ξ è una funzione intera, dunque più "perfetta" dal punto di vista anche solo dell'estetica, inoltre possiede due rappresentazioni sotto forma di serie e di prodotto infinito viste in precedenza.
Porre

$$\xi\left(\frac{1}{2}+it\right), \qquad t \in \mathbb{C},$$

nelle intenzioni del suo autore ha sicuramente l'obiettivo di dire, in maniera "elegante", che tutti gli zeri sono reali.
Magari, in termini *moderni*, si sarebbe detto

$$\xi\left(\frac{1}{2}+is\right), \qquad s \in \mathbb{C},$$

poiché in seguito con la "t" si sarebbe indicata $Im(s)$ invece che una variabile complessa come fa Riemann nel suo articolo; tuttavia non cambia quanto detto fino ad ora.
L'ipotesi di Riemann universalmente accettata dice, infatti, che "tutti gli zeri non banali della funzione ζ hanno parte reale $1/2$", mentre l'originale afferma che "è ragionevole supporre che gli zeri (della ξ) siano reali". C'è una questione puramente estetica che, nelle intenzioni del matematico tedesco, rende più elegante un'affermazione che tratta di "zeri reali" piuttosto che una che ha come oggetto "zeri con parte reale $1/2$". Gli assi cartesiani sono elementi fondamentali del piano complesso e sono il riferimento per la rappresentazione stessa dei complessi (sia trigonometrica che algebrica).
Tali accorgimenti estetici, inoltre, hanno anche ripercussioni su risultati successivi. Riemann, infatti, parla di

$$\left(T\log\frac{T}{2\pi}-T\right)i,$$

mentre noi troveremo un

$$\frac{T}{2\pi}\log\frac{T}{2\pi}-\frac{T}{2\pi}$$

come risultato dell'integrale citato dallo stesso matematico nel suo articolo.

14.1.2 Stima per la ζ

Richiamiamo due risultati già visti nella sezione dedicata alla Teoria Analitica dei Numeri (§10.2.3). Il primo è la rappresentazione, ampiamente utilizzata anche in seguito, che si ricava dalla formula di somma di Eulero

$$\zeta(s) = -\frac{1}{1-s} - s\int_1^{+\infty} \frac{t-\lfloor t\rfloor}{t^{s+1}}dt + 1 = \frac{s}{s-1} - s\int_1^{+\infty} \frac{t-\lfloor t\rfloor}{t^{s+1}}dt,$$

mentre il secondo è la somma parziale che ci conduce proprio a tale risultato (leggermente riadattata e senza utilizzo della notazione O grande)

$$\sum_{n\le x}\frac{1}{n^s} = \frac{x^{1-s}-1}{1-s} - s\int_1^{x} \frac{t-\lfloor t\rfloor}{t^{s+1}}dt + 1 - \frac{x-\lfloor x\rfloor}{x^s}.$$

In quest'ultima, facendo qualche calcolo (a parte) si ottiene

$$\frac{x^{1-s}-1}{1-s} + 1 = \frac{1}{(1-s)x^{s-1}} - \frac{1}{1-s} + 1 = -\frac{1}{(s-1)x^{s-1}} - \frac{s}{1-s}$$
$$= -\frac{1}{(s-1)x^{s-1}} + \frac{s}{s-1},$$

che, nella formula iniziale, ci dà

$$\sum_{n\le x}\frac{1}{n^s} = -\frac{1}{(s-1)x^{s-1}} + \frac{s}{s-1} - s\int_1^{x} \frac{t-\lfloor t\rfloor}{t^{s+1}}dt - \frac{x-\lfloor x\rfloor}{x^s}.$$

In generale c'è tutta una serie di formulazioni equivalenti per questa somma parziale che si possono ottenere l'una dall'altra operando somme intermedie simili a quelle viste per trovare quest'ultimo risultato: ci serviremo di questo perché semplificherà i calcoli nella dimostrazione del seguente teorema.

Teorema ([5], §6)

Sia $s = \sigma + it$, con $t \ge 1$. Allora, per quanto riguarda la funzione ζ, valgono le seguenti stime:

$$|\zeta(s)| < A\log(t), \qquad \sigma \ge 1,$$

inoltre, con $0 < \delta < 1$,

$$|\zeta(s)| < A(\delta)t^{1-\delta}, \qquad \sigma \ge \delta.$$

In esse A è una costante opportuna (nel secondo caso dipendente da δ).

Dimostrazione

Ricordiamo le due rappresentazioni della ζ richiamate poco fa:

$$\zeta(s) = -\frac{1}{1-s} - s\int_1^{+\infty} \frac{t-\lfloor t\rfloor}{t^{s+1}}dt + 1$$

e quella della somma parziale x −esima

$$\sum_{n\le x}\frac{1}{n^s} = -\frac{1}{(s-1)x^{s-1}} + \frac{s}{s-1} - s\int_1^{x} \frac{t-\lfloor t\rfloor}{t^{s+1}}dt - \frac{x-\lfloor x\rfloor}{x^s}$$

che valgono per $x \ge 1, s \ne 1$ tale che $\sigma > 0$.

Sottraendo alla ζ la sua somma parziale, abbiamo, per $\sigma > 0, t \ge 1, x \ge 1$

$$\zeta(s) - \sum_{n \leq x} \frac{1}{n^s}$$
$$= \frac{s}{s-1} - s\int_1^{+\infty} \frac{t - \lfloor t \rfloor}{t^{s+1}} dt + \frac{1}{(s-1)x^{s-1}} - \frac{s}{s-1} + s\int_1^{x} \frac{t - \lfloor t \rfloor}{t^{s+1}} dt$$
$$+ \frac{x - \lfloor x \rfloor}{x^s} = -s\int_x^{\infty} \frac{t - \lfloor t \rfloor}{t^{s+1}} dt + \frac{1}{(s-1)x^{s-1}} + \frac{x - \lfloor x \rfloor}{x^s}.$$

Portando la somma parziale al secondo membro e passando al modulo otteniamo la stima

$$|\zeta(s)| \leq \left| \sum_{n \leq x} \frac{1}{n^s} \right| + |s| \left| \int_x^{\infty} \frac{t - \lfloor t \rfloor}{t^{s+1}} dt \right| + \left| \frac{1}{(s-1)x^{s-1}} \right| + \left| \frac{x - \lfloor x \rfloor}{x^s} \right|,$$

ricordando che $x - \lfloor x \rfloor \leq 1$ e che $|s| = |\sigma + it| \leq \sigma + t$, e inoltre

$$|s| \left| \int_x^{\infty} \frac{t - \lfloor t \rfloor}{t^{s+1}} dt \right| \leq |s| \int_x^{\infty} \frac{dx}{x^{\sigma+1}} = |s| \left. \frac{x^{-\sigma}}{-\sigma} \right|_x^{\infty} = \frac{|s|}{\sigma x^{\sigma}} \leq \left(1 + \frac{t}{\sigma}\right) \frac{1}{x^{\sigma}},$$

possiamo calcolare i termini di quella stima ottenendo

$$|\zeta(s)| \leq \sum_{n \leq x} \frac{1}{n^{\sigma}} + |s| \int_x^{\infty} \frac{dx}{x^{\sigma+1}} + \frac{1}{tx^{\sigma-1}} + \frac{1}{x^{\sigma}} \leq \sum_{n \leq x} \frac{1}{n^{\sigma}} + \left(1 + \frac{t}{\sigma}\right) \frac{1}{x^{\sigma}} + \frac{1}{tx^{\sigma-1}} + \frac{1}{x^{\sigma}}.$$

A questo punto distinguiamo due casi.

Per $\sigma \geq 1$, ricordando che $x \geq 1$,

$$|\zeta(s)| \leq \sum_{n \leq x} \frac{1}{n^{\sigma}} + \left(1 + \frac{t}{\sigma}\right) \frac{1}{x^{\sigma}} + \frac{1}{tx^{\sigma-1}} + \frac{1}{x^{\sigma}} \leq \sum_{n \leq x} \frac{1}{n} + \frac{1+t}{x} + \frac{1}{t} + \frac{1}{x}$$
$$\leq \log(x+1) + 3 + \frac{t}{x},$$

nella quale si è sfruttata la seguente disuguaglianza

$$\sum_{n \leq x} \frac{1}{n} \leq \log(x),$$

oltre al fatto che

$$\frac{1}{x} \leq 1, \qquad \frac{1}{t} \leq 1,$$

che vale poiché $x \geq 1$, $t \geq 1$.

In questa formula, ponendo $x = t$ otteniamo

$$|\zeta(s)| \leq \log(t+1) + 4 < A \log(t),$$

per A costante opportuna, il che dimostra la prima tesi.

Passiamo al caso $\sigma \geq \delta$, con $0 < \delta < 1$, per dimostrare la seconda implicazione del teorema.

Il ragionamento è analogo al precedente, si fa la differenza tra la ζ e la sua somma parziale per poi isolare la ζ stessa e passare ai moduli

$$|\zeta(s)| \leq \left| \sum_{n \leq x} \frac{1}{n^s} \right| + |s| \left| \int_x^{\infty} \frac{t - \lfloor t \rfloor}{t^{s+1}} dt \right| + \left| \frac{1}{(s-1)x^{s-1}} \right| + \left| \frac{x - \lfloor x \rfloor}{x^s} \right|$$

da cui, proprio come prima,

$$|\zeta(s)| \leq \sum_{n \leq x} \frac{1}{n^{\sigma}} + \left(1 + \frac{t}{\sigma}\right) \frac{1}{x^{\sigma}} + \frac{1}{tx^{\sigma-1}} + \frac{1}{x^{\sigma}}.$$

Tuttavia, possiamo andare oltre poiché sappiamo che

$$\frac{1}{x^\sigma} \leq \frac{1}{x^\delta}, \qquad \frac{t}{\sigma} \leq \frac{t}{\delta}, \qquad (t \geq 1)$$

dunque

$$|\zeta(s)| \leq \sum_{n \leq x} \frac{1}{n^\sigma} + \left(1 + \frac{t}{\sigma}\right)\frac{1}{x^\sigma} + \frac{1}{tx^{\sigma-1}} + \frac{1}{x^\sigma} \leq \sum_{n \leq x} \frac{1}{n^\delta} + \left(1 + \frac{t}{\delta}\right)\frac{1}{x^\delta} + \frac{1}{tx^{\delta-1}} + \frac{1}{x^\delta}$$

$$\leq \int_0^{\lfloor x \rfloor} \frac{dx}{x^\delta} + \frac{x^{1-\delta}}{t} + \frac{3t}{\delta x^\delta} \leq \frac{x^{1-\delta}}{1-\delta} + x^{1-\delta} + \frac{2t}{\delta x^\delta}.$$

Ora, ponendo $x = t$, abbiamo

$$|\zeta(s)| \leq t^{1-\delta}\left(\frac{1}{1-\delta} + 1 + \frac{3}{\delta}\right) = A(\delta)t^{1-\delta}, \qquad \sigma \geq \delta, \qquad t \geq 1,$$

che dimostra la seconda implicazione, poiché A è un parametro che varia con δ.

Per quanto riguarda la prima implicazione, si era visto che

$$|\zeta(s)| \leq \log(t+1) + 4 < A\log(t),$$

dove A è un'opportuna costante e $t \geq 1$. Tuttavia, essa si può anche esprimere nel modo seguente

$$|\zeta(s)| = O(\log(t)), \qquad t \geq 1, \qquad \sigma \geq 1,$$

la quale fornisce una relazione asintotica per la ζ.
Sarà utile, inoltre, enunciare il seguente lemma.

Lemma ([28], §6.4)
Per T sufficientemente grande, se sul segmento $-1 < \sigma \leq 2$, con $t = T$ non vi sono zeri della funzione ζ, allora si ha

$$\frac{\zeta'(\sigma + iT)}{\zeta(\sigma + iT)} = \sum_{\substack{\rho \\ |T-\gamma|<1}} \frac{1}{\sigma + iT - \rho} + O(\log(T)).$$

Inoltre il numero degli addendi della somma è $O(\log(T))$.

La sommatoria di questo teorema è calcolata lungo gli zeri che si trovano "vicino" (si dice, infatti, $|T - \gamma| < 1$) al valore immaginario T lungo il segmento $]-1 + iT, 2 + iT]$.
La dimostrazione di questo teorema è un lungo calcolo che trae le sue fondamenta dalla derivata logaritmica della ζ (intesa nel modo che vedremo più avanti, parlando della formula esplicita). Si tratta, infatti, di isolare la derivata logaritmica della ζ presente nella formula che riguarda quella della funzione ξ come vedremo nella sottosezione dedicata alla formula esplicita per la ψ.

14.1.3 Il principio dell'argomento

Inizieremo con il mostrare un risultato sulle funzioni meromorfe – cioè analitiche tranne che per singolarità di tipo polo (§3.4.3) – che prende il nome di principio dell'argomento.

Teorema (Principio dell'argomento) ([29], §6.6)
Sia γ una curva semplice e chiusa e contenuta in un dominio $\Omega \subseteq \mathbb{C}$. Sia f una funzione analitica in D eccetto un numero finito di poli nella regione racchiusa da γ. Si supponga, inoltre, che $f(s) \neq 0$ per $s \in \gamma$, allora

$$\frac{1}{2\pi i}\int_{\gamma}\frac{f'(s)}{f(s)}ds = N_0 - N_p,$$

nella quale N_0 e N_p sono, rispettivamente, gli zeri e i poli della funzione $f(s)$ all'interno di γ contati con la loro molteplicità.

Dimostrazione
Indichiamo, senza troppe precisazioni tecniche, la regione racchiusa dal cammino γ con la scrittura $Int(\gamma)$.
Iniziamo con l'osservare che, poiché per ipotesi gli zeri e i poli di f sono all'interno della curva γ, allora l'integrando

$$\frac{f'(s)}{f(s)}$$

è analitico nel bordo del cammino (cioè per $s \in \gamma$).
Le singolarità dell'integrando possono incorrere solo nei casi di zeri o poli per la $f(s)$ proprio perché nei restanti punti di $Int(\gamma)$ questa funzione è analitica, dunque non ci sono singolarità di sorta in quel rapporto.
Per dimostrare il teorema inizieremo trattando il singolo caso di uno zero e di un polo per poi estendere il discorso al caso generale.
Supponiamo, dunque, che $s_0 \in Int(\gamma)$ sia uno zero di ordine n per f, dunque possiamo scrivere

$$f(s) = (s - s_0)^n g(s), \qquad n \geq 1,$$

nella quale $g(s)$ è una funzione analitica in s_0 e $g(s_0) \neq 0$ in accordo a quanto detto per gli zeri di una funzione nella sezione dedicata ai richiami di analisi complessa (§3.4.2).
Da questa scrittura possiamo, dunque, calcolare $f'(s)$ con le usuali regole di derivazione

$$f'(s) = (s - s_0)^n g'(s) + n(s - s_0)^{n-1} g(s),$$

per poi dividere tale espressione per f.
Dunque, in un qualche disco bucato $\widetilde{D}(s_0, R)$ – cioè $D(s_0, R)$ senza il punto s_0 – con R tale che tale disco sia contenuto in $Int(\gamma)$, si ha

$$\frac{f'(s)}{f(s)} = \frac{(s - s_0)^n g'(s) + n(s - s_0)^{n-1} g(s)}{(s - s_0)^n g(s)} = \frac{g'(s)}{g(s)} + \frac{n}{s - s_0}.$$

Questo risultato ci mostra che l'integrando ha un polo semplice per $s = s_0$ (in quanto $g(s_0) \neq 0$ per costruzione). In tale polo, il residuo (§3.5.1) è

$$Res\left(\frac{f'(s)}{f(s)}, s_0\right) = \lim_{s \to s_0}\left[(s - s_0)\left(\frac{g'(s)}{g(s)} + \frac{n}{s - s_0}\right)\right] = \lim_{s \to s_0}\left(\frac{g'(s)(s - s_0)}{g(s)} + n\right) = n,$$

che, per costruzione, è l'ordine dello zero di f in s.
Possiamo, ora, fare un analogo ragionamento indicando con $s_p \in Int(\gamma)$ un polo di ordine m per la funzione $f(s)$ ottenendo, in $\widetilde{D}(s_p, R')$,

$$\frac{f'(s)}{f(s)} = \frac{(s - s_p)^{-m} h'(s) - m(s - s_p)^{-m-1} h(s)}{(s - s_p)^{-m} h(s)} = \frac{h'(s)}{h(s)} + \frac{-m}{s - s_p},$$

che si basa sul fatto che se f ha un polo di ordine m in s_p essa si può scrivere (§3.4.3) come

$$f(s) = \frac{h(s)}{(s - s_p)^m}, \qquad m \geq 1,$$

dove $h(s)$ è olomorfa in s_p.

Dalla scrittura precedente vediamo che l'integrando $f'(s)/f(s)$ in s_p ha un polo semplice: se ne calcoliamo il residuo, troviamo che esso è uguale a $-m$ che è l'opposto dell'ordine del polo di f in s_0.

A questo punto possiamo concludere la dimostrazione estendendo i ragionamenti al caso generale.

Se, infatti, $s_{0_1}, s_{0_2}, \ldots, s_{0_r} \in \backslash Int(\gamma)$ sono gli zeri di f con molteplicità $n_1, n_2, \ldots, n_r$ e $s_{p_1}, s_{p_2}, \ldots, s_{p_q}$ sono i poli di f di ordine rispettivamente $m_1, m_2, \ldots, m_s$, allora troveremo che i residui di tali zeri e tali poli sono, rispettivamente, $n_1, n_2, \ldots, n_r$ per gli zeri e $-m_1, -m_2, \ldots, -m_s$ per i poli. Quindi, per il teorema dei residui (§3.5.1)

$$\int_\gamma \frac{f'(s)}{f(s)} ds = 2\pi i \left[\sum_{k=1}^{r} Res\left(\frac{f'(s)}{f(s)}, s_{0_k}\right) + \sum_{h=1}^{q} Res\left(\frac{f'(s)}{f(s)}, s_{p_h}\right) \right]$$

$$= 2\pi i \left[\sum_{k=1}^{r} n_k + \sum_{h=1}^{q} (-m_k) \right] = 2\pi i (N_0 - N_p).$$

In essa, dividere ambo i membri per $2\pi i$ ci consente di ottenere la tesi del teorema.

Questo teorema è anche chiamato "teorema dell'indicatore logaritmico" poiché possiamo notare che l'integrando è la derivata logaritmica di $f(s)$. Parleremo di derivata logaritmica nella sottosezione dedicata alla formula esplicita per la funzione ψ di Chebyshev dimostrata da von Mangoldt.

Il nome "principio dell'argomento" deriva dal numero di volte che l'argomento di f cambia quando s varia lungo il cammino di integrazione.

Per capire meglio questo concetto, consideriamo la funzione $f(s) = 1/s^2$ e come curva γ la circonferenza unitaria $|s| = 1$. All'interno di γ la funzione f è analitica tranne un polo di ordine 2 nell'origine (come si può constatare dalla scrittura stessa della funzione).

Possiamo parametrizzare $\gamma = e^{it}, 0 \leq t \leq 2\pi$.

Quindi, $f(s)$ calcolata lungo il cerchio unitario così parametrizzato diventa

$$f(\gamma(t)) = \frac{1}{e^{2it}}.$$

Sappiamo (§3.2.10) che

$$\arg(e^{2it}) = \arg(e^{it} \cdot e^{it}) = \arg(e^{it}) + \arg(e^{it}) = 2\arg(e^{it}),$$

quindi l'argomento di f varia due volte – cioè arriva da 0 a 4π – percorrendo la circonferenza unitaria così parametrizzata.

Con la seguente notazione, cioè

$$[\arg(e^{2it})]_\gamma,$$

indichiamo la variazione dell'argomento di e^{2it} lungo la curva γ che, in base a quanto detto, è 4π.

14.1.4 La densità degli zeri

In questo paragrafo forniremo una dimostrazione del teorema di Riemann-von Mangoldt circa la densità degli zeri.
Iniziamo con l'introdurre la quantità

$$N(T) = \text{numero degli zeri } \rho \text{ della } \xi, \text{con } 0 < Re(\rho) < 1, 0 < Im(\rho) \leq T.$$

Generalmente non si considerano gli zeri ρ con $Im(\rho) < 0$, in quanto $\xi(s) = \xi(\bar{s})$. Grazie a questa proprietà, possiamo notare che gli zeri ρ della funzione ξ con $-T \leq Im(\rho) \leq T$ e $0 < Re(\rho) < 1$ sono $2N(T)$.
Come abbiamo visto ad inizio paragrafo, nel suo articolo di ricerca Riemann dà la stima che, in seguito, verrà provata da von Mangoldt circa mezzo secolo dopo. Ci sono diverse dimostrazioni riguardo tale risultato, noi proporremo quella di Backlund (1914) ([5], §9; [28], §6.4).

Teorema
Per $T \to +\infty$, vale la seguente formula

$$N(T) = \frac{T}{2\pi}\log\left(\frac{T}{2\pi}\right) - \frac{T}{2\pi} + O(\log(T)).$$

Dimostrazione
Supponiamo di avere $T > 3$ tale che $s = \sigma + iT$ soddisfa $\xi(s) \neq 0$, cioè T non sia la parte immaginaria di uno zero della funzione ξ.
Consideriamo, ora, il rettangolo R come mostra la Figura 14.1.

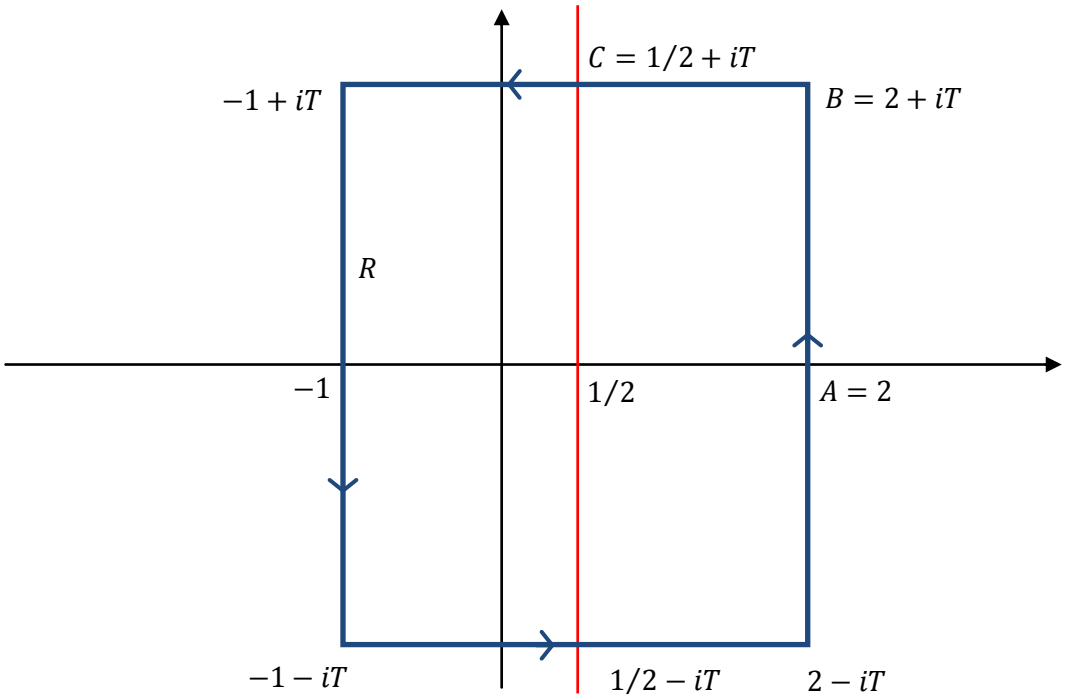

Figura 14.1. Rettangolo di integrazione per $\xi'(s)/\xi(s)$.

In Figura 14.1 in rosso è rappresentata la linea critica: da notare che essa è la mediana di due lati opposti del rettangolo. Infatti R è "centrato" in modo da contenere la striscia critica al suo interno (dato che è quella la regione dove sono contenuti gli zeri).
Consideriamo, dunque, R.
Come già detto, in R la funzione ξ ha $2N(T)$ zeri, in quanto $\xi(\bar{s}) = \xi(s)$. Per la scelta di T, inoltre, $\xi(s)$ non ha zeri nel bordo di tale rettangolo.
Ora, per il principio dell'argomento o, meglio, per l'osservazione successiva su tale principio

$$2N(T) = \frac{1}{2\pi}\left[\arg\big(\xi(s)\big)\right]_R,$$

in essa, la notazione $\left[\arg\big(\xi(s)\big)\right]_R$ denota la variazione di $\arg\big(\xi(s)\big)$ lungo il bordo di R, come detto prima di enunciare questo teorema.
Utilizzeremo una notazione analoga per dimostrare la crescita dell'argomento di tutti i termini.
Ricordiamo la definizione di $\xi(s)$ a partire dall'equazione funzionale della ζ (§13.1.1)

$$\xi(s) = (s-1)\Gamma\left(\frac{s}{2}+1\right)\zeta(s)\pi^{-\frac{s}{2}} = \frac{1}{2}s(s-1)\Gamma\left(\frac{s}{2}\right)\zeta(s)\pi^{-\frac{s}{2}}.$$

In essa, abbiamo posto, sfruttando le proprietà della funzione Γ come estensione del fattoriale naturale (§8.2)

$$\Gamma\left(\frac{s}{2}+1\right) = \frac{s}{2}\Gamma\left(\frac{s}{2}\right),$$

poiché una tale sostituzione permetterà una semplificazione successiva nei calcoli in quanto $\Gamma(s/2)$ è più facile da trattare rispetto a $\Gamma(s/2+1)$.
Sfruttiamo, ora, le proprietà dell'argomento

$$\left[\arg\big(\xi(s)\big)\right]_R = \left[\arg\left(\frac{1}{2}s(s-1)\right)\right]_R + \left[\arg\big(\eta(s)\big)\right]_R,$$

nella quale, si è posto (per comodità)

$$\eta(s) = \pi^{-\frac{s}{2}}\Gamma\left(\frac{s}{2}\right)\zeta(s);$$

per come è definita, si può notare che $\eta(s) = \eta(1-s)$.
Ora, si ha

$$\left[\arg\left(\frac{1}{2}s(s-1)\right)\right]_R = 4\pi,$$

in quanto, lungo il rettangolo, il fattore s^2 cambia il suo argomento 2 volte: infatti, per il principio dell'argomento visto in precedenza, posto

$$g(s) = \frac{1}{2}s(s-1),$$

risulta

$$\int_R \frac{g'(s)}{g(s)}ds = 4\pi i,$$

poiché $g(s)$ ha solo 2 zeri interni a R e nessun polo.
Ora, dal fatto che $\eta(s) = \eta(1-s)$ e $\eta(s) = \eta(\bar{s})$ – proprietà che deriva direttamente dalla ξ –, sfruttando tutte queste simmetrie

$$\begin{aligned}\left[\arg(\eta(s))\right]_R &= 4\left[\arg(\eta(s))\right]_{ABC}\\ &= 4\left[\arg\left(\pi^{-\frac{s}{2}}\right)\right]_{ABC} + 4\left[\arg\left(\Gamma\left(\frac{s}{2}\right)\right)\right]_{ABC} + 4\left[\arg(\zeta(s))\right]_{ABC},\end{aligned}$$

in quanto, come si può vedere in Figura 14.1, ABC equivale a $1/4$ del totale del rettangolo.
Ricapitoliamo, dunque, la situazione

$$2N(T) = \frac{1}{2\pi}\left[\arg(\xi(s))\right]_R,$$

e, per quanto visto fino ad ora si ottiene

$$2N(T) = \frac{1}{2\pi}\left(4\pi + 4\left[\arg\left(\pi^{-\frac{s}{2}}\right)\right]_{ABC} + 4\left[\arg\left(\Gamma\left(\frac{s}{2}\right)\right)\right]_{ABC} + 4\left[\arg(\zeta(s))\right]_{ABC}\right),$$

da cui

$$\pi N(T) = \pi + \arg\left(\pi^{-\frac{s}{2}}\right) + \left[\arg\left(\Gamma\left(\frac{s}{2}\right)\right)\right]_{ABC} + \left[\arg(\zeta(s))\right]_{ABC}.$$

A questo punto, poiché l'argomento di $f(s)$ è la parte immaginaria di $\log(f(s))$ (§3.2.10), abbiamo

$$\left[\arg\left(\pi^{-\frac{s}{2}}\right)\right]_{ABC} = \left[-\frac{t}{2}\log(\pi)\right]_{ABC} = -\frac{T}{2}\log(\pi).$$

Per quanto riguarda la funzione Γ, ricordiamo dalla formula di Stirling generalizzata (§8.3) che, per $\delta > 0$, si ha

$$\log(\Gamma(z)) = \left(z - \frac{1}{2}\right)\log(z) - z + \frac{1}{2}\log(2\pi) + O(|z|^{-1}),$$

per $|z| \to +\infty$ nell'angolo $|\arg(z)| \leq \pi - \delta$.
Possiamo, dunque, utilizzarla per il calcolo dell'argomento della funzione Γ

$$\begin{aligned}\left[\arg\left(\Gamma\left(\frac{s}{2}\right)\right)\right]_{ABC} &= \left[Im\left(\log\left(\Gamma\left(\frac{s}{2}\right)\right)\right)\right]_{ABC} = Im\left(\log\left(\Gamma\left(\frac{1}{4} + \frac{1}{2}iT\right)\right)\right) - Im(\log(\Gamma(1)))\\ &= Im\left[\left(-\frac{1}{4} + \frac{1}{2}iT\right)\log\left(\frac{1}{2}iT\right) - \frac{1}{2}iT + \frac{1}{2}\log(2\pi) + O(T^{-1})\right]\\ &= \frac{1}{2}T\log\left(\frac{1}{2}T\right) - \frac{1}{8}\pi - \frac{T}{2} + O(T^{-1}), \qquad T \to +\infty.\end{aligned}$$

Ora, riunendo tutte le formule trovate fino ad ora, si ha

$$N(T) = \frac{T}{2\pi}\log\left(\frac{T}{2\pi}\right) - \frac{T}{2\pi} + \frac{7}{8} + \frac{1}{\pi}\left[\arg(\zeta(s))\right]_{ABC} + O\left(\frac{1}{T}\right).$$

Resta solo da dimostrare che $\left[\arg(\zeta(s))\right]_{ABC} = O(\log(T))$.
Innanzitutto, ricordiamo che $\zeta(2) \in \mathbb{R}$, proprio dal fatto che la funzione ζ stessa è definita (per $Re(s) > 1$) come estensione della serie armonica generalizzata.
Sappiamo, inoltre, che $\zeta(2 + it) \neq 0$ e

$$Re(\zeta(2 + it)) \geq 1 - \sum_{n=2}^{\infty}\frac{1}{n^2} > 1 - \frac{1}{2^2} - \int_2^{\infty}\frac{du}{u^2} = \frac{1}{4},$$

quindi $Re(\zeta(2 + it)) > 0$, dunque, da questo abbiamo

$$\left|\arg(\zeta(2 + iT))\right| < \frac{\pi}{2},$$

in altre parole la variazione dell'argomento sul tratto verticale è limitata.

Resta solo da considerare la variazione sul tratto orizzontale BC. Ricordiamo che, per il principio dell'argomento, abbiamo

$$\left[\arg\left(\zeta(s)\right)\right]_{BC} = \int_{\frac{1}{2}}^{2} Im\left(\frac{\zeta'(\sigma+iT)}{\zeta(\sigma+iT)}\right) d\sigma.$$

Ci serve, quindi, una stima per l'argomento dell'integrale lungo il segmento $1/2 + iT$ e $2 + iT$.

Il lemma precedente implica che

$$\begin{aligned}\left[\arg\left(\zeta(s)\right)\right]_{BC} &= -\int_{\frac{1}{2}}^{2} Im\left(\frac{\zeta'(\sigma+iT)}{\zeta(\sigma+iT)}\right) d\sigma + O(1)\\ &= -\sum_{\substack{\rho\\ |T-\gamma|<1}} \int_{\frac{1}{2}}^{2} Im\left(\frac{1}{\sigma+iT-\rho}\right) d\sigma + O(\log(T))\\ &= \sum_{\substack{\rho\\ |T-\gamma|<1}} [\arg(\sigma+iT-\rho)]_{BC} + O(\log(T)).\end{aligned}$$

Ora, se $Re\left(\zeta(s)\right)$ si annullasse q volte tra $2 + iT$ e $\frac{1}{2} + iT$, questo intervallo sarebbe diviso in $q + 1$ parti, in ognuna delle quali $Re\left(\zeta(s)\right) \geq 0$ oppure $Re\left(\zeta(s)\right) \leq 0$.

In ognuna di queste parti, la variazione di $\arg\left(\zeta(s)\right)$ non sarebbe più grande di π (ricordiamo che l'argomento in sé è un angolo). Quindi, sempre per il lemma, sappiamo che quella sommatoria ha, al massimo, $O(\log(T))$ addendi.

Ma allora, unendo i due risultati, abbiamo

$$\left[\arg\left(\zeta(s)\right)\right]_{ABC} = \pi + \pi O(\log(T)) = O(\log(T)).$$

A questo punto, inserendo il risultato nella formula trovata in precedenza

$$N(T) = \frac{T}{2\pi}\log\left(\frac{T}{2\pi}\right) - \frac{T}{2\pi} + \frac{7}{8} + \frac{1}{\pi}\left[\arg\left(\zeta(s)\right)\right]_{ABC} + O\left(\frac{1}{T}\right),$$

si ottiene la tesi del teorema

$$N(T) = \frac{T}{2\pi}\log\left(\frac{T}{2\pi}\right) - \frac{T}{2\pi} + \frac{7}{8} + \frac{1}{\pi}O(\log(T)) + O\left(\frac{1}{T}\right) = \frac{T}{2\pi}\log\left(\frac{T}{2\pi}\right) - \frac{T}{2\pi} + O(\log(T)),$$

quest'ultimo inglobando la costante nell'O grande e tenendo conto che

$$\frac{7}{8} + O(\log(T)) + O\left(\frac{1}{T}\right) = O(\log(T)),$$

poiché $1/T$ e $7/8$ sono di ordine inferiore a $\log(T)$.

14.2 LA FORMULA DI PERRON

In questa sottosezione dimostreremo la così detta formula di Perron che sarà il tassello fondamentale per la successiva dimostrazione di von Mangoldt circa il collegamento tra la funzione ψ e gli zeri (non banali) della ζ.

14.2.1 Alcune proprietà preliminari

Ricapitoliamo brevemente alcune proprietà degli integrali che utilizzeremo in questa sezione. Supponiamo di avere f, una funzione analitica in $\Omega \subseteq \mathbb{C}$ e γ una curva Ω.

(i) Innanzitutto ricordiamo (§3.3.2)

$$\int_{\gamma} f(s)ds = -\int_{-\gamma} f(s)ds.$$

(ii) Analogamente agli integrali reali

$$\int_{\gamma} |f(s)|ds \leq \int_{\gamma} Mds\,, \qquad M \geq |f(s)|, \qquad s \in \gamma.$$

Possiamo, infatti, minorare il modulo di una funzione con il suo massimo lungo il cammino di integrazione proprio come accade per il confronto tra integrali reali (proprietà utilizzata soprattutto per vedere la convergenza di integrali impropri).
La si era accennata – un po' in secondo piano – nella sezione di analisi complessa (§3.3.2).

(iii) Ricordiamo che, se $|f(s)| \leq M$

$$\left|\int_{\gamma} f(s)ds\right| \leq \int_{\gamma} |f(s)||ds| \leq M \cdot l(\gamma),$$

in cui

$$\int_{\gamma} |ds| = \int_{a}^{b} |\gamma'(t)|dt = l(\gamma),$$

dove $l(\gamma)$ è la lunghezza della curva.
In particolare, vediamo l'esempio del seguente integrale:

$$\int_{a-ih}^{K-ih} |ds| = \int_{a}^{K} |\gamma'(t)|dt = \int_{a}^{K} dt.$$

Se, infatti, andassimo a parametrizzare il segmento di estremi $[a - ih, K - ih]$, avremmo

$$\gamma(t) = t - ih, \qquad t \in [a, K],$$

da cui

$$\gamma'(t) = 1,$$

che spiega l'ultimo passaggio.

La dimostrazione della formula di Perron passa attraverso la stima di vari integrali per la quale utilizzeremo ampiamente queste proprietà appena ricordate.

14.2.2 Valutazione di un integrale

In questo paragrafo seguiremo Edwards ([9], §3.3) per procedere alla valutazione del seguente integrale

$$\frac{1}{2\pi}\int_{a-i\infty}^{a+\infty}\frac{x^s}{s}ds = \lim_{h\to\infty}\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{x^s}{s}ds\,, \qquad a>0, \qquad x>0.$$

Il calcolo si basa sulle tecniche di integrazione lungo curve dell'analisi complessa e, soprattutto, del teorema dei residui (§3.5.1). Distingueremo 3 casi:

- $0 < x < 1$,
- $x = 1$,
- $x > 1$.

Un'osservazione piuttosto banale, ma utile, è che l'integrando, essendo un quoziente tra funzioni intere ha una sola singolarità (di tipo polo) in $s = 0$ che annulla il denominatore.

Caso $0 < x < 1$

Il procedimento, analogamente agli altri casi, è quello di rifarsi al calcolo di integrali lungo curve e, poi, di portarsi al limite per ottenere il risultato che si vuole. In questo caso consideriamo il rettangolo – che chiameremo R – che racchiude la regione

$$\{a \le Re(s) \le K, -h \le Im(s) \le h\},$$

in cui $K > a$ è una costante arbitrariamente grande.

Per avere un raffronto grafico, si può considerare l'immagine in Figura 14.2 che, tra l'altro, segnala anche l'ubicazione della singolarità dell'integrando che è esterna al rettangolo in questione.

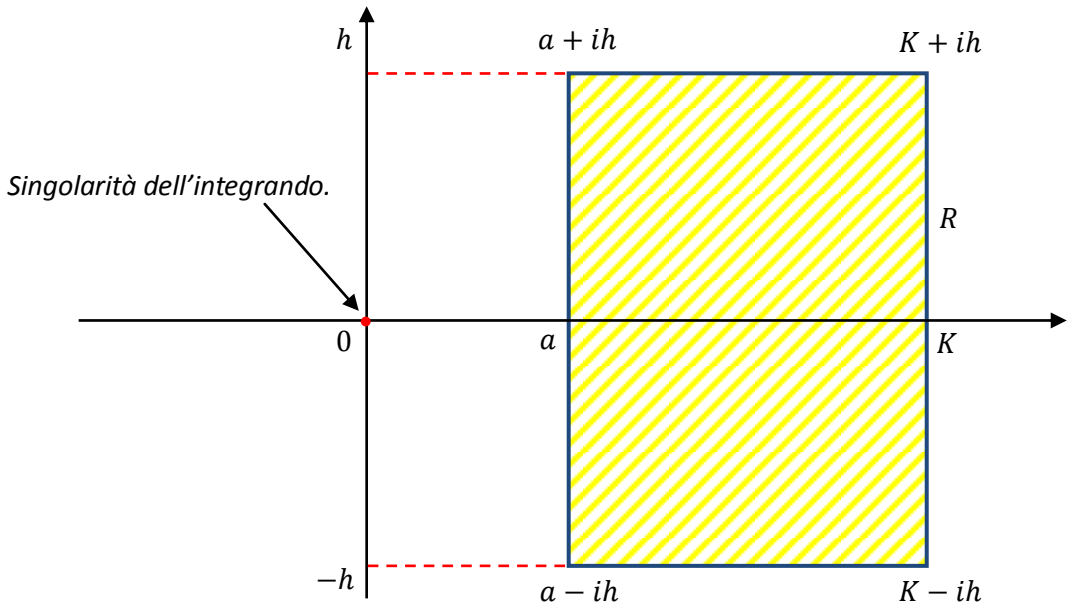

Figura 14.2. Rettangolo di integrazione.

Poiché all'interno del rettangolo l'integrando non ha singolarità, l'integrale lungo il rettangolo è nullo (§3.3.4) per vari teoremi visti nell'integrazione complessa.

Quindi, prendendo come verso di percorrenza quello antiorario,

$$\frac{1}{2\pi i}\int_R \frac{x^s}{s}ds = \frac{1}{2\pi i}\int_{a+ih}^{a-ih}\frac{x^s}{s}ds + \frac{1}{2\pi i}\int_{a-ih}^{K-ih}\frac{x^s}{s}ds + \frac{1}{2\pi i}\int_{K-ih}^{K+ih}\frac{x^s}{s}ds + \frac{1}{2\pi i}\int_{K+ih}^{a+ih}\frac{x^s}{s}ds$$
$$= 0,$$

proprio per quanto appena detto.

Ci interessa, nel nostro caso, il primo di questa serie di integrali che è proprio quello che vogliamo stimare. Isoliamolo, dunque, rispetto agli altri portandolo all'altro membro

$$\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{x^s}{s}ds=\frac{1}{2\pi i}\int_{a-ih}^{K-ih}\frac{x^s}{s}ds+\frac{1}{2\pi i}\int_{K-ih}^{K+ih}\frac{x^s}{s}ds-\frac{1}{2\pi i}\int_{a+ih}^{K+ih}\frac{x^s}{s}ds.$$

Pensiamo dunque a trattare i tre integrali al secondo membro separatamente.
Iniziamo con il primo, cioè

$$\frac{1}{2\pi i}\int_{K-ih}^{K+ih}\frac{x^s ds}{s}.$$

Diamo, dunque, una stima sfruttando le proprietà (ii) e (iii) viste ad inizio sottosezione.

$$\left|\frac{1}{2\pi i}\int_{K-ih}^{K+ih}\frac{x^s}{s}ds\right|=\frac{1}{2\pi}\left|\int_{K-ih}^{K+ih}\frac{x^s}{s}ds\right|\leq\frac{1}{2\pi}\int_{K-ih}^{K+ih}\left|\frac{x^s}{s}\right||ds|=\frac{1}{2\pi}\int_{K-ih}^{K+ih}\frac{x^K}{K}|ds|$$
$$=\frac{1}{2\pi}\frac{x^K}{K}\int_{K-ih}^{K+ih}|ds|=2h\frac{x^K}{2\pi K},$$

che tende a zero per $K\to\infty$, in quanto $0<x<1$.
La curva in questione è il segmento che unisce i due punti $(K-ih)$ e $(K+ih)$, dunque è la distanza tra i punti stessi, cioè $2h$.
Gli altri due integrali sono molto simili e, con lo stesso procedimento – leggermente riadattato – si riesce ad avere una stima in entrambi i casi.
Ne analizzeremo solo uno e, cioè,

$$\frac{1}{2\pi}\int_{a-ih}^{K-ih}\frac{x^s ds}{s},$$

poiché l'altro caso è analogo a questo. Procediamo, dunque, con la stima, servendoci delle proprietà (ii) e (iii) viste in precedenza

$$\left|\frac{1}{2\pi i}\int_{a-ih}^{K-ih}\frac{x^s}{s}ds\right|=\frac{1}{2\pi}\left|\int_{a-ih}^{K-ih}\frac{x^s}{s}ds\right|\leq\frac{1}{2\pi}\int_{a-ih}^{K-ih}\frac{|x^s|}{|s|}|ds|\leq\frac{1}{2\pi h}\int_{a-ih}^{K-ih}|x^s||ds|$$
$$=\frac{1}{2\pi h}\int_a^K x^t dt=\frac{1}{2\pi h}\left|\frac{x^t}{\log(x)}\right|_a^K\Bigg|=\frac{1}{2\pi}\frac{|x^K-x^a|}{h|\log(x)|}.$$

In essa si è applicata la proprietà (ii) richiamata in precedenza sfruttando il fatto che

$$\frac{1}{|s|}\leq\frac{1}{h},$$

per $s\in[a-ih,K-ih]$: cosa che si può provare facilmente parametrizzando il segmento come mostrato nella (iii). Quest'ultima, invece, è entrata in gioco nel calcolo dell'integrale sulla lunghezza d'arco.
Per l'altro la dimostrazione è analoga, con un'altrettanto analoga conclusione.
Otteniamo passando ai moduli

$$\left|\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{x^s}{s}ds\right|\leq 2\cdot\frac{1}{2\pi}\frac{|x^K-x^a|}{h|\log(x)|}+\frac{x^K}{2\pi K},$$

nel quale il secondo membro tende a 0 per $h,K\to+\infty$ e si arriva alla conclusione cercata.

Caso $x=1$

In questo caso, invece di passare per il teorema dei residui, ci si serve del calcolo diretto. A tal proposito richiamiamo il seguente risultato

$$\int_{-\infty}^{+\infty} \frac{1}{1+x^2} dx = \tan^{-1}(x)|_{-\infty}^{\infty} = \pi.$$

Iniziamo con il calcolo nel caso finito, per poi passare al limite per $h \to +\infty$ servendoci dell'integrale appena richiamato e delle proprietà enunciate ad inizio sottosezione, nel paragrafo precedente. Si ha

$$\frac{1}{2\pi i}\int_{a-ih}^{a+ih} \frac{ds}{s} = \frac{1}{2\pi i}\int_{-h}^{h} \frac{idt}{a+it} = \frac{1}{2\pi}\int_{-h}^{h} \frac{(a-it)dt}{(a-it)(a+it)} = \frac{1}{2\pi}\int_{-h}^{h} \frac{a-it}{a^2+t^2} dt$$
$$= \frac{1}{2\pi}\int_{-h}^{h} \frac{adt}{a^2+t^2} = \frac{1}{2\pi}\int_{-h}^{h} \frac{a}{1/a^2\,(1+t^2/a^2)} dt = \frac{1}{2\pi}\int_{-h/a}^{h/a} \frac{du}{1+u^2},$$

dove nell'ultimo passaggio si è operato un semplice cambio di variabile $t/a = u$, da cui $du = dt/a$. Inoltre

$$\frac{1}{2\pi}\int_{-h}^{h} \frac{a-it}{a^2+t^2} dt = \frac{1}{2\pi}\int_{-h}^{h} \frac{a}{a^2+t^2} dt - i\frac{1}{2\pi}\int_{-h}^{h} \frac{t}{a^2+t^2} dt = \frac{1}{2\pi}\int_{-h}^{h} \frac{a}{a^2+t^2} dt,$$

in quanto il secondo integrale è nullo poiché è un'integrale di una funzione dispari lungo un dominio simmetrico all'origine.

Passiamo, dunque, al limite

$$\lim_{h\to+\infty} \frac{1}{2\pi i}\int_{a-ih}^{a+ih} \frac{ds}{s} = \lim_{h\to+\infty} \frac{1}{2\pi}\int_{-h/a}^{h/a} \frac{du}{1+u^2} = \frac{1}{2},$$

risultato che si ottiene sfruttando l'integrale visto in precedenza.

Caso $x > 1$

In questo caso ci serviremo di un percorso di integrazione – non molto differente da quello in Figura 14.2 – per valutare l'integrale.

Si consideri, dunque, il rettangolo (che chiameremo nuovamente R) che racchiude la seguente regione di piano complesso, per $K > a$ arbitrariamente grande

$$\{-K \le Re(s) \le a, -h \le Im(s) \le h\}.$$

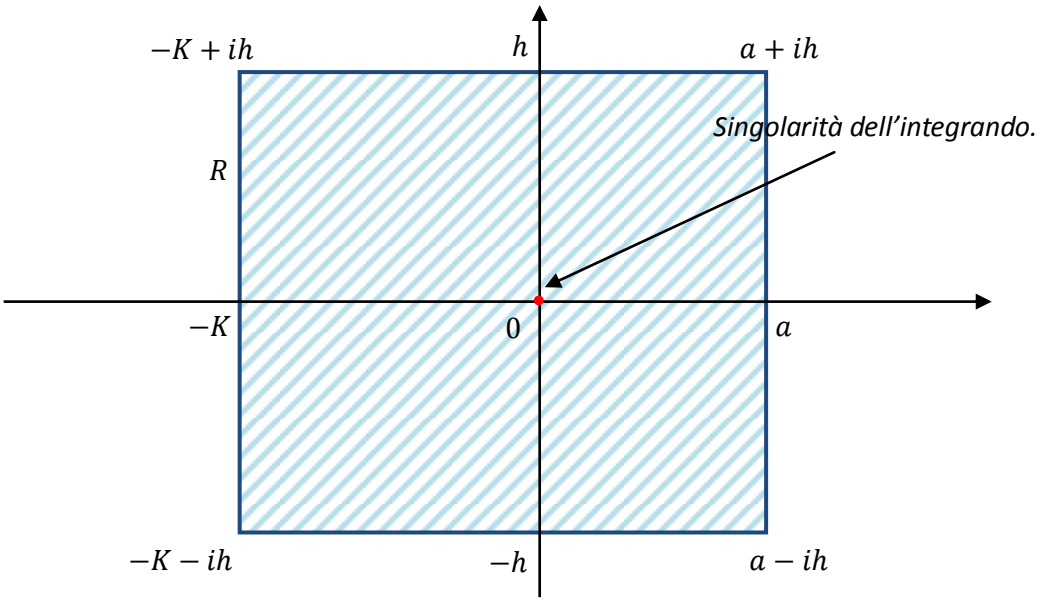

Figura 14.3. Rettangolo di integrazione.

Come si può vedere in Figura 14.3, stavolta la singolarità nell'origine è interna al dominio di integrazione, dunque

$$\frac{1}{2\pi i}\int_R \frac{x^s}{s}ds = Res\left(\frac{x^s}{s}, 0\right),$$

per il teorema dei residui (§3.5.1), tenendo conto che l'indice di avvolgimento è 1 e che il cammino di integrazione è percorso in senso antiorario.
Calcoliamo allora

$$Res\left(\frac{x^s}{s}, 0\right) = \lim_{s\to 0}\left(\frac{x^s}{s}\cdot s\right) = 1,$$

sfruttando la formula per il calcolo del residuo.
A questo punto possiamo operare un calcolo simile a quello visto nel caso $0 < x < 1$, tenendo conto, però, che ora $x > 1$.

$$\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{x^s ds}{s} + \frac{1}{2\pi i}\int_{a+ih}^{-K+ih}\frac{x^s ds}{s} + \frac{1}{2\pi i}\int_{-K+ih}^{-K-ih}\frac{x^s ds}{s} + \frac{1}{2\pi i}\int_{-K+ih}^{a-ih}\frac{x^s ds}{s} = 1.$$

In essa, vogliamo stimare

$$\left|\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{x^s ds}{s} - 1\right|,$$

che si ottiene portando il primo termine all'altro membro e ricordando che $|-z| = |z|$.

$$\begin{aligned}\left|\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{x^s ds}{s} - 1\right| &\le \left|\frac{1}{2\pi i}\int_{a+ih}^{-K+ih}\frac{x^s ds}{s}\right| + \left|\frac{1}{2\pi i}\int_{-K+ih}^{-K-ih}\frac{x^s ds}{s}\right| + \left|\frac{1}{2\pi i}\int_{-K+ih}^{a-ih}\frac{x^s ds}{s}\right| \\ &\le \frac{1}{2\pi h}\int_{-K}^{a} x^\sigma d\sigma + \frac{2h}{2\pi}\cdot\frac{x^{-K}}{K} + \frac{1}{2\pi h}\int_{-K}^{a} x^\sigma d\sigma = \frac{1}{\pi}\frac{x^a - x^{-K}}{h\log(x)} + \frac{1}{\pi}\frac{x^{-K}h}{K},\end{aligned}$$

con procedimenti analoghi al caso $0 < x < 1$, ma con l'unica differenza che, per $x > 1$, $\log(x) > 0$ e $x^\sigma - x^{-K} > 0$, dunque non necessitano del modulo.
Ora, portando $K \to +\infty$, si ottiene

$$\left|\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{x^s ds}{s} - 1\right| \le \frac{1}{\pi}\frac{x^a}{h\log(x)}.$$

A questo punto, tenendo conto che $a > 0$ e $x > 1$, per $h \to +\infty$ si ottiene la stima cercata, cioè

$$\left|\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{x^s ds}{s} - 1\right| \to 0,$$

e quindi

$$\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\frac{x^s}{s}ds = \lim_{h\to\infty}\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{x^s}{s}ds = 1, \qquad a > 1, \qquad x > 1.$$

Conclusione
Valutando quest'integrale nei tre casi proposti abbiamo dimostrato il seguente

Teorema (Formula di Perron) ([28], §6.5)
Per $x > 0$ e $a > 0$, si ha

$$\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\frac{x^s ds}{s} = \begin{cases} 0, & x \in]0,1[, \\ \frac{1}{2}, & x = 1, \\ 1, & x > 1. \end{cases}$$

Questo risultato possiede molte varianti. L'importante è che al posto della " x " nell'integrando, possiamo inserire qualsiasi argomento "a valori reali" ottenendo analoghi risultati che si rifanno al precedente. A tal proposito, possiamo enunciare il seguente corollario.

Corollario ([30])
Per $x > 0, a > 0$ e n intero positivo

$$\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\left(\frac{x}{n}\right)^s\frac{ds}{s} = \begin{cases} 0, & n > x \\ \frac{1}{2}, & n = x, \\ 1, & n < x. \end{cases}$$

14.3 LA FORMULA ESPLICITA PER LA ψ

La formula di Perron è un tassello fondamentale che ci consente di dimostrare la formula di Riemann-von Mangoldt.

14.3.1 Il legame tra la ψ e la ζ

In questo paragrafo vedremo un risultato sulla funzione ζ che è alla base della formula di Riemann-von Mangoldt.
Richiamiamo la seguente serie di uguaglianze (§11.5), valida per $Re(s) > 1$

$$\frac{\zeta'(s)}{\zeta(s)} = \frac{1}{\zeta(s)}\frac{d}{ds}\left(\prod_{p \text{ primo}}\frac{1}{1-p^{-s}}\right) = \frac{1}{\zeta(s)}\sum_{p \text{ primo}}\frac{d}{dz}\left(\frac{1}{1-p^{-s}}\right)\prod_{q\neq p \text{ primo}}\frac{1}{1-q^{-s}}$$

$$= \sum_{p \text{ primo}}\left(\frac{-p^{-s}\log(p)}{(1-p^{-s})^2}\right)(1-p^{-s}) = \sum_{p \text{ primo}} -\frac{\log(p)}{p^s}\left(1-\frac{1}{p^s}\right)^{-1}$$

$$= -\sum_{p \text{ primo}}\log(p)\sum_{m=1}^{\infty}\frac{1}{p^{ms}} = -\sum_{n=2}^{\infty}\frac{\Lambda(n)}{n^s},$$

che prova il fatto che

$$\frac{\zeta'(s)}{\zeta(s)} = -\sum_{p \text{ primo}}\log(p)\sum_{m=1}^{\infty}\frac{1}{p^{ms}} = -\sum_{n=2}^{\infty}\frac{\Lambda(n)}{n^s}.$$

Analizziamo ora il seguente integrale servendoci della precedente formula di Perron

$$\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} -\frac{\zeta'(s)}{\zeta(s)}x^s\frac{ds}{s} = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\sum_{p \text{ primo}}\log(p)\sum_{m=1}^{\infty}\frac{1}{p^{ms}}x^s\frac{ds}{s}.$$

Ricaviamo

$$\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\sum_{p\text{ primo}}\log(p)\sum_{m=1}^{\infty}\frac{1}{p^{ms}}x^s\frac{ds}{s}=\sum_{p}\log(p)\sum_{m\geq 1}\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\left(\frac{x}{p^{-m}}\right)^s\frac{ds}{s}.$$

La formula di Perron ci serve per valutare l'ultima formula, tenendo conto del fatto che p è primo.

Supponiamo di avere $x>1$ non intero, in modo da evitare automaticamente il caso $x=p^m$ all'interno di tale formula.

- Se $x>p^m$, quell'integrale vale 1.
- Se $x<p^m$, quell'integrale è nullo.

Quindi, in quella sommatoria restano solo i termini tali che $p^m<x$:

$$\sum_{p}\log(p)\sum_{m\geq 1}\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\left(\frac{x}{p^{-m}}\right)^s\frac{ds}{s}=\sum_{\substack{m\geq 1\\ p:p^m<x}}\log(p)=\psi(x).$$

Ricordiamo, infatti, la definizione della funzione di Chebyshev (§10.2.4)

$$\psi(x)=\sum_{n\leq x}\Lambda(n)=\sum_{\substack{m=1\\ p^m\leq x}}^{\infty}\sum_{p\text{ primo}}\Lambda(p^m)=\sum_{m=1}^{\infty}\sum_{p\leq x^{1/m}}\log(p),$$

con la differenza che, stavolta, il $\leq$ non è contemplato poiché abbiamo considerato x non intero. Quest'espressione, tuttavia, si può estendere a qualunque x con accorgimenti tecnici dei quali non abbiamo trattato.

Tornando all'integrale iniziale, la relazione trovata ci dà un legame tra la ζ e la ψ:

$$\psi(x)=\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}-\frac{\zeta'(s)}{\zeta(s)}x^s\frac{ds}{s},\qquad Re(s)>1,\qquad x>1,$$

la cui valutazione analitica porterà alla formula esplicita per la ψ.

14.3.2 Qualche considerazione sulla convergenza

Abbiamo, dunque, ottenuto

$$\psi(x)=\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}-\frac{\zeta'(s)}{\zeta(s)}x^s\frac{ds}{s},\qquad Re(s)>1,\qquad x>1.$$

Tuttavia dobbiamo ancora valutare la convergenza di tale integrale e vedere se è possibile calcolarlo come somma termine a termine (passando dall'integrale della somma alla somma degli integrali):

$$\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\sum_{p\text{ primo}}\log(p)\sum_{m=1}^{\infty}\frac{1}{p^{ms}}x^s\frac{ds}{s}=\sum_{p}\log(p)\sum_{m\geq 1}\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\left(\frac{x}{p^{-m}}\right)^s\frac{ds}{s}.$$

In seguito vedremo anche la convergenza dei singoli termini che verranno fuori nella dimostrazione della formula esplicita di von Mangoldt per la ψ. Pensiamo, per ora, a chiarire brevemente questo punto, ripartendo dalla serie di uguaglianze richiamata nel paragrafo precedente

$$\sum_{p \text{ primo}} \log(p) \sum_{m=1}^{\infty} \frac{1}{p^{ms}} = \sum_{n=2}^{\infty} \frac{\Lambda(n)}{n^s}.$$

Nel primo integrale, dunque,

$$\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} \sum_{p \text{ primo}} \log(p) \sum_{m=1}^{\infty} \frac{1}{p^{ms}} x^s \frac{ds}{s} = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} \sum_{n=2}^{\infty} \frac{\Lambda(n)}{n^s} x^s \frac{ds}{s}$$

$$= \sum_{n=2}^{\infty} \Lambda(n) \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} \left(\frac{x}{n}\right)^s \frac{ds}{s} = \lim_{h\to+\infty}\left(\sum_{n=2}^{\infty} \Lambda(n) \frac{1}{2\pi i}\int_{a-ih}^{a+ih} \left(\frac{x}{n}\right)^s \frac{ds}{s}\right).$$

Valuteremo solo i termini $n > x$ poiché i termini $n < x$ sono, comunque, finiti in quanto $x > 1$ è fissato. Nella dimostrazione della formula di Perron, si è provata, nel caso $0 < x < 1$ (quindi quella del nostro caso poiché $0 < x/n < 1$, per $n > x$), la seguente stima

$$\left|\frac{1}{2\pi i}\int_{a-ih}^{K-ih} \frac{x^s}{s} ds\right| = \frac{1}{2\pi} \frac{|x^K - x^a|}{h|\log(x)|}.$$

Applicandola, con i dovuti accorgimenti, al termine in esame, ne otteniamo una simile

$$\left|\Lambda(n) \frac{1}{2\pi i}\int_{a-ih}^{a+ih} \left(\frac{x}{n}\right)^s \frac{ds}{s}\right| \le \log(n) \frac{x^a}{n^a \pi h(\log(n) - \log(x))} \le cost \cdot \frac{1}{n^a h};$$

per $h \to +\infty$ questi singoli termini tendono a zero (in modulo), così abbiamo provato che il limite può essere calcolato termine a termine.

14.3.3 La "pericolosità" del logaritmo complesso e l'arte di "differenziare logaritmicamente"

Soffermiamoci brevemente sul logaritmo complesso introdotto in (§3.2.10). Per $w \in \mathbb{C}\backslash\{0\}$

$$\log w = \ln|\mathrm{w}| + i(\theta + 2k\pi).$$

In molti testi è possibile anche trovare la seguente scrittura

$$\lg(w) = \log|w| + i(\theta + 2k\pi),$$

nella quale il logaritmo complesso di un $w \in \mathbb{C}\backslash\{0\}$ è indicato con $\lg(w)$ mentre con $\log|w|$ è indicato l'usuale logaritmo reale del numero reale $|w|$ (ricordiamo che il modulo di un complesso è una quantità reale non negativa).

Tuttavia, a parte questi dettagli puramente estetici, per il logaritmo complesso non valgono proprietà a cui siamo abituati con il logaritmo reale, come l'iniettività dello stesso.

Vediamo di fare un esempio pratico.

Se volessimo risolvere la (semplice) equazione

$$e^x = 1,$$

applicheremmo il logaritmo (reale) ad entrambi i membri ottenendo, grazie alla suddetta iniettività,

$$\log(e^x) = \log(e^0), \qquad x = 0.$$

Passiamo ora al campo complesso, considerando l'analoga equazione

$$e^z = 1.$$

Stavolta però il logaritmo non è definito univocamente come la funzione inversa dell'esponenziale e anzi si ha:

- $\log(e^z) = \ln|e^z| + 2k\pi i = 2k\pi i, k \in \mathbb{Z}$;
- $\log 1 = 2k\pi i, k \in \mathbb{Z}$.

Il logaritmo complesso ha dunque "infiniti" rami regolari. Il discorso si generalizza ad equazioni

$$f(z) = g(z),$$

per le quali il passaggio al logaritmo complesso porta a

$$\log\big(f(z)\big) = \log\big(g(z)\big) + 2k\pi i,$$

per k intero. Tuttavia, derivando ambo i membri, si ha l'uguaglianza

$$\frac{f'(z)}{f(z)} = \frac{g'(z)}{g(z)},$$

priva di ogni ambiguità (basta moltiplicare ambo i membri per $f(z)$ o $g(z)$ per ottenere in entrambi la derivata dell'equazione di partenza).

Abbiamo mostrato le basi teoriche per le quali è lecito applicare l'operazione comunemente chiamata "derivata logaritmica di una funzione $f(z)$ " ad entrambi i membri di un'uguaglianza.

Data una funzione $f(z)$ a valori complessi, definiremo allora la sua derivata logaritmica [4] come

$$\big(\log\big(f(z)\big)\big)' = \frac{f'(z)}{f(z)}.$$

Il suo impiego consente di trattare più agevolmente delle uguaglianze tra funzioni di variabile complessa, soprattutto nel caso di prodotti (anche infiniti).

Per esempio, prendendo la formula prodotto per la funzione ξ , che dimostreremo nell'Appendice IV,

$$\xi(s) = \xi(0)\prod_{\rho}\left(1 - \frac{s}{\rho}\right),$$

otteniamo

$$\log\big(\xi(s)\big) = \log\big(\xi(0)\big) + \sum_{\rho}\log\left(1 - \frac{s}{\rho}\right),$$

da cui, derivando, giungiamo alla derivata logaritmica

$$\frac{\xi'(s)}{\xi(s)} = \sum_{\rho}\left(-\frac{1}{\rho}\right)\left(\frac{1}{1 - s/\rho}\right).$$

14.3.4 Formula esplicita: base

Dobbiamo valutare il seguente integrale

$$\psi(x) = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} -\frac{\zeta'(s)}{\zeta(s)}x^s\frac{ds}{s}, \qquad Re(s) > 1, \qquad x > 1,$$

che, come già detto, racchiude in sé il legame tra la funzione ζ e la ψ. In questo modo arriveremo a quella che è comunemente chiamata "formula esplicita" (per la ψ) o "formula/teorema di Riemann-von Mangoldt".
Si tratta di eguagliare le due scritture della funzione ξ – cioè quella derivante dall'equazione funzionale della ζ e la formula prodotto – per poi derivare logaritmicamente ambo i membri di tale uguaglianza.
Per semplicità di calcoli ci serviremo della scrittura originale di Riemann con la funzione $\prod$ invece della funzione Γ (§8.4).
Ricordiamo, dunque, la definizione della funzione ξ

$$\xi(s) = (s-1)\Gamma\left(\frac{s}{2}+1\right)\zeta(s)\pi^{-\frac{s}{2}},$$

che si traduce in

$$\xi(s) = (s-1)\prod\left(\frac{s}{2}\right)\zeta(s)\pi^{-\frac{s}{2}},$$

utilizzando la scrittura originale di Riemann ($\prod(s) = \Gamma(s+1)$). D'altro lato abbiamo

$$\xi(s) = \xi(0)\prod_{\rho}\left(1-\frac{s}{\rho}\right).$$

Possiamo, dunque, eguagliare le due scritture, ottenendo

$$\xi(0)\prod_{\rho}\left(1-\frac{s}{\rho}\right) = (s-1)\prod\left(\frac{s}{2}\right)\zeta(s)\pi^{-\frac{s}{2}}.$$

Iniziamo prendendo in considerazione il primo membro.
La derivata logaritmica della formula prodotto per la funzione ξ l'avevamo vista alla fine del paragrafo precedente ottenendo

$$\frac{\xi'(s)}{\xi(s)} = \sum_{\rho}\left(-\frac{1}{\rho}\right)\left(\frac{1}{1-s/\rho}\right).$$

Passiamo, dunque, al secondo membro.
Ci serviamo della funzione $\prod$ proprio perché ha una forma più semplice, rispetto alla Γ:

$$\prod(s) = \prod_{n=1}^{\infty}\frac{(1+1/n)^s}{1+s/n}, \qquad s \in \mathbb{C},$$

vista nella sezione dedicata, appunto, alla funzione Γ.
Calcolata in $s/2$, si ottiene

$$\prod\left(\frac{s}{2}\right) = \prod_{n=1}^{\infty}\frac{(1+1/n)^{s/2}}{1+s/(2n)}$$

Calcoliamo, a parte, la derivata logaritmica della funzione $\prod$, iniziando proprio con il logaritmo.

$$\log\left(\prod\left(\frac{s}{2}\right)\right) = \log\left(\prod_{n=1}^{\infty}\frac{(1+1/n)^{s/2}}{1+s/(2n)}\right) = \sum_{n=1}^{\infty}\left(\log\left(1+\frac{1}{n}\right)^{s/2} - \log\left(1+\frac{s}{2n}\right)\right)$$
$$= \sum_{n=1}^{\infty}\left(\frac{s}{2}\log\left(1+\frac{1}{n}\right) - \log(2n+s) + \log(2n)\right).$$

Passando, dunque, alla derivata logaritmica otteniamo

$$\frac{\left(\prod(s/2)\right)'}{\prod(s/2)} = \sum_{n=1}^{\infty}\left(\frac{1}{2}\log\left(1+\frac{1}{n}\right)-\frac{1}{2n+s}\right),$$

che andremo, in seguito, a sostituire quando otterremo la derivata logaritmica del secondo membro dell'uguaglianza iniziale.

Come abbiamo detto, l'utilità della derivata logaritmica sta anche nel fatto che consente di scomporre prodotti, anche complicati, sfruttando le proprietà del logaritmo – che "trasforma prodotti in somme" – e, in seguito, le ben note proprietà di linearità della derivata.

Possiamo, dunque, derivare logaritmicamente il secondo membro dell'uguaglianza iniziale. Iniziamo proprio con il logaritmo ricordando che $\pi^{-s/2} = e^{-\frac{s}{2}\log(\pi)}$:

$$\begin{aligned}\log\big(\xi(s)\big) &= \log(s-1)+\log\left(\prod\left(\frac{s}{2}\right)\right)+\log\big(\zeta(s)\big)+\log\left(e^{-\frac{s}{2}\log(\pi)}\right)\\ &= \log(s-1)+\log\left(\prod\left(\frac{s}{2}\right)\right)+\log\big(\zeta(s)\big)-\frac{s}{2}\log(\pi).\end{aligned}$$

Arriviamo, infine, alla derivata logaritmica, riportando quanto detto per la funzione $\prod$:

$$\frac{\xi'(s)}{\xi(s)} = \frac{1}{s-1}+\sum_{n=1}^{\infty}\left(\frac{1}{2}\log\left(1+\frac{1}{n}\right)-\frac{1}{2n+s}\right)+\frac{\zeta'(s)}{\zeta(s)}-\frac{1}{2}\log(\pi).$$

A questo punto resta solamente da eguagliare ambo i membri

$$\sum_{\rho}\left(-\frac{1}{\rho}\right)\left(\frac{1}{1-s/\rho}\right) = \frac{1}{s-1}+\sum_{n=1}^{\infty}\left(\frac{1}{2}\log\left(1+\frac{1}{n}\right)-\frac{1}{2n+s}\right)+\frac{\zeta'(s)}{\zeta(s)}-\frac{1}{2}\log(\pi),$$

tuttavia, prima di operare una qualsiasi semplificazione, ponendo $s=0$, si ottiene

$$\sum_{\rho}\left(-\frac{1}{\rho}\right) = -1+\sum_{n=1}^{\infty}\left(\frac{1}{2}\log\left(1+\frac{1}{n}\right)-\frac{1}{2n}\right)+\frac{\zeta'(0)}{\zeta(0)}-\frac{1}{2}\log(\pi).$$

Sottraiamo, ora, membro a membro quest'ultima uguaglianza alla precedente

$$\begin{aligned}&\sum_{\rho}\left(-\frac{1}{\rho}\right)\left(\frac{1}{1-s/\rho}\right)-\sum_{\rho}\left(-\frac{1}{\rho}\right)\\ &\qquad= \frac{1}{s-1}+\sum_{n=1}^{\infty}\left(\frac{1}{2}\log\left(1+\frac{1}{n}\right)-\frac{1}{2n+s}\right)+\frac{\zeta'(s)}{\zeta(s)}-\frac{1}{2}\log(\pi)+1\\ &\qquad\quad-\sum_{n=1}^{\infty}\left(\frac{1}{2}\log\left(1+\frac{1}{n}\right)-\frac{1}{2n}\right)-\frac{\zeta'(0)}{\zeta(0)}+\frac{1}{2}\log(\pi).\end{aligned}$$

In quest'ultima espressione compare $\zeta'(s)/\zeta(s)$, il termine che dovremo isolare per poi sostituirlo nell'integrale che vogliamo valutare. Per ora, però, procediamo con il semplificare membro a membro questa *definitiva* uguaglianza così ottenuta.

(i) Il primo è piuttosto semplice e non necessita di molti calcoli.

$$\begin{aligned}\sum_{\rho}\left(-\frac{1}{\rho}\right)\left(\frac{1}{1-s/\rho}\right)-\sum_{\rho}\left(-\frac{1}{\rho}\right) &= \sum_{\rho}\left(-\frac{1}{\rho}\right)\left(\frac{1}{1-s/\rho}-1\right)\\ &= \sum_{\rho}\left(-\frac{1}{\rho}\right)\left(\frac{\rho}{\rho-s}-1\right) = \sum_{\rho}\left(-\frac{1}{\rho}\right)\left(\frac{s}{\rho-s}\right) = \sum_{\rho}\frac{s}{\rho(s-\rho)}.\end{aligned}$$

(ii) Nel secondo, prima di tutto esaminiamo la differenza tra le sommatorie con gli stessi indici, servendoci delle proprietà delle stesse (§1.2.2)

$$\sum_{n=1}^{\infty}\left(\frac{1}{2}\log\left(1+\frac{1}{n}\right)-\frac{1}{2n+s}\right)-\sum_{n=1}^{\infty}\left(\frac{1}{2}\log\left(1+\frac{1}{n}\right)-\frac{1}{2n}\right)$$
$$=\sum_{n=1}^{\infty}\left(-\frac{1}{2n+s}+\frac{1}{2n}\right).$$

Infine

$$\frac{1}{s-1}+\sum_{n=1}^{\infty}\left(-\frac{1}{2n+s}+\frac{1}{2n}\right)+\frac{\zeta'(s)}{\zeta(s)}-\frac{1}{2}\log(\pi)+1-\frac{\zeta'(0)}{\zeta(0)}+\frac{1}{2}\log(\pi)$$
$$=\frac{s}{s-1}+\sum_{n=1}^{\infty}\frac{s}{2n(2n+s)}+\frac{\zeta'(s)}{\zeta(s)}-\frac{\zeta'(0)}{\zeta(0)}.$$

Nell'uguaglianza, dunque, otteniamo

$$\sum_{\rho}\frac{s}{\rho(s-\rho)}=\frac{s}{s-1}+\sum_{n=1}^{\infty}\frac{s}{2n(2n+s)}+\frac{\zeta'(s)}{\zeta(s)}-\frac{\zeta'(0)}{\zeta(0)}.$$

Possiamo, dunque, isolare il termine $\zeta'(s)/\zeta(s)$ (mantenendo il segno negativo che si ottiene portandolo al primo membro)

$$-\frac{\zeta'(s)}{\zeta(s)}=\frac{s}{s-1}-\sum_{\rho}\frac{s}{\rho(s-\rho)}+\sum_{n=1}^{\infty}\frac{s}{2n(2n+s)}-\frac{\zeta'(0)}{\zeta(0)},$$

per poi ottenere, nell'integrale iniziale

$$\psi(x)=\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}-\frac{\zeta'(s)}{\zeta(s)}x^s\frac{ds}{s}$$
$$=\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\frac{s}{s-1}x^s\frac{ds}{s}-\sum_{\rho}\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\frac{s}{\rho(s-\rho)}x^s\frac{ds}{s}$$
$$+\sum_{n=1}^{\infty}\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\frac{s}{2n(2n+s)}x^s\frac{ds}{s}-\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\frac{\zeta'(0)}{\zeta(0)}x^s\frac{ds}{s}.$$

Lo valuteremo, anche qui, membro a membro servendoci della formula di Perron.

Prima di cominciare, però, dovremmo chiederci se una sostituzione del genere sia lecita, analizzando la convergenza dei singoli membri.

- La serie

$$\sum_{n=1}^{\infty}\frac{s}{2n(2n+s)},$$

  è uniformemente convergente per s fissato (è un $O(n^2)$).

- La serie

$$\sum_{\rho}\frac{\rho}{\rho(s-\rho)}=\sum_{\rho}\frac{1}{s-\rho},$$

  è più difficile da valutare.

  Degli zeri non banali della funzione ζ si sa soltanto che sono simmetrici rispetto alla retta $Re(s)=1/2$ (al massimo giacciono sulla stessa). Ad ogni zero ρ, ne corrisponde un analogo $1-\rho$ simmetrico rispetto alla già citata retta $Re(s)=1/2$.

  Per ogni zero, dunque, abbiamo (per $s\in\mathbb{C}$ fissato)

$$\left|\frac{1}{s-\rho}+\frac{1}{s-(1-\rho)}\right| = \left|\frac{1}{(s-1/2)-(\rho-1/2)}+\frac{1}{(s-1/2)+(\rho-1/2)}\right|$$
$$= \left|\frac{(s-1/2)+(\rho-1/2)-(s-1/2)+(\rho-1/2)}{(s-1/2)^2-(\rho-1/2)^2}\right|$$
$$= \left|\frac{2(s-1/2)}{(s-1/2)^2-(\rho-1/2)^2}\right| = Cost\cdot\left|\rho-\frac{1}{2}\right|^{-2}.$$

Quest'ultima proprio perché s è una quantità complessa (finita) fissata.
Ora dovremmo dimostrare la convergenza della serie

$$\sum_{\rho}\frac{1}{|\rho-1/2|^{-2}},$$

cosa che non faremo (ora) in quanto è un argomento ampiamente trattato nell'Appendice IV dedicato alla formula prodotto di Hadamard per la ξ.
La convergenza di questa serie, infatti, è un passo fondamentale proprio per la convergenza di tale prodotto.

14.3.5 Formula esplicita: dimostrazione

L'obiettivo del paragrafo precedente è stato quello di fornire la base della dimostrazione della formula esplicita per la ψ. Abbiamo, infatti, trattato una serie di passaggi intermedi atti a verificare, oltre che la sostituzione della formula per la ψ ottenuta nel legame tra questa e la ζ, anche la liceità della stessa finendo per concludere con un "sì" poiché i singoli termini di tale sostituzione convergono (anche uniformemente) per s complesso.
Abbiamo, dunque

$$\psi(x) = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} -\frac{\zeta'(s)}{\zeta(s)}x^s\frac{ds}{s}$$
$$= \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\frac{s}{s-1}x^s\frac{ds}{s} - \sum_{\rho}\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\frac{s}{\rho(s-\rho)}x^s\frac{ds}{s}$$
$$+\sum_{n=1}^{\infty}\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\frac{s}{2n(2n+s)}x^s\frac{ds}{s} - \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\frac{\zeta'(0)}{\zeta(0)}x^s\frac{ds}{s}.$$

Per i singoli termini, abbiamo dimostrato la convergenza (anche uniforme) per s complesso fissato e, in generale, per $|s| \le K$. Dovremmo – in linea teorica – passare attraverso il limite in quanto non sappiamo nulla per $|s| \to \infty$.
La dimostrazione "formalmente corretta", dovrebbe essere la seguente

$$\psi(x) = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} -\frac{\zeta'(s)}{\zeta(s)} x^s \frac{ds}{s}$$

$$= \lim_{h\to+\infty}\left(\frac{1}{2\pi i}\int_{a-ih}^{a+ih} \frac{s}{s-1} x^s \frac{ds}{s} - \sum_{\rho} \frac{1}{2\pi i}\int_{a-ih}^{a+ih} \frac{s}{\rho(s-\rho)} x^s \frac{ds}{s}\right.$$

$$\left. + \sum_{n=1}^{\infty} \frac{1}{2\pi i}\int_{a-ih}^{a+ih} \frac{s}{2n(2n+s)} x^s \frac{ds}{s} - \frac{1}{2\pi i}\int_{a-ih}^{a+ih} \frac{\zeta'(0)}{\zeta(0)} x^s \frac{ds}{s}\right)$$

$$= \lim_{h\to+\infty}\left(\frac{1}{2\pi i}\int_{a-ih}^{a+ih} \frac{s}{s-1} x^s \frac{ds}{s}\right) - \lim_{h\to+\infty}\left(\sum_{\rho} \frac{1}{2\pi i}\int_{a-ih}^{a+ih} \frac{s}{\rho(s-\rho)} x^s \frac{ds}{s}\right)$$

$$+ \lim_{h\to+\infty}\left(\sum_{n=1}^{\infty} \frac{1}{2\pi i}\int_{a-ih}^{a+ih} \frac{s}{2n(2n+s)} x^s \frac{ds}{s}\right) - \lim_{h\to+\infty}\left(\frac{1}{2\pi i}\int_{a-ih}^{a+ih} \frac{\zeta'(0)}{\zeta(0)} x^s \frac{ds}{s}\right).$$

La convergenza dei termini, inoltre, dimostra anche che è lecito suddividere il limite di quella somma come somma dei singoli limiti.
Analizzeremo separatamente i primi tre integrali (su quattro) ottenuti nell'ultima espressione per poi concludere separatamente con l'ultimo di questi.

I primi tre integrali

Per quanto riguarda i primi tre integrali, essi – a meno di costanti moltiplicative (che si possono portare fuori dall'integrale) – sono tutti del tipo

$$\frac{1}{2\pi i}\int_{a-ih}^{a+ih} \frac{x^s}{s-\beta} ds,$$

nel quale β è un'opportuna costante (complessa) variabile a seconda dell'integrale stesso.
Poniamo, dunque, $t = s - \beta$, con $\beta = b + ic$ da cui $dt = ds$,

$$\frac{1}{2\pi i}\int_{a-ih}^{a+ih} \frac{x^s}{s-\beta} ds = \frac{1}{2\pi i}\int_{(a-b)-i(h-c)}^{(a-b)+i(h-c)} \frac{x^{t+\beta}}{t} dt,$$

che non è più un'integrale con estremi simmetrici rispetto all'asse reale come nel caso precedente.
Vogliamo rapportarci al seguente integrale

$$\frac{1}{2\pi i}\int_{(a-b)-ih}^{(a+b)+ih} \frac{x^{t+\beta}}{t} dt,$$

che è simmetrico rispetto all'asse reale e consente di applicare nuovamente le formule di Perron. Proviamo ad immaginare la situazione. Abbiamo quattro punti nel piano complesso:

- $A = (a-b) - ih$;
- $A' = (a-b) - i(h-c)$;
- $B = (a-b) + ih$;
- $B' = (a-b) + i(h-c)$.

Tutto questo ragionamento vale perché a può essere un qualsiasi reale ($a > 0$ per servirci della formula di Perron): dunque b può essere un qualsiasi altro reale, basta che, poi, scegliamo a tale che $a - b > 0$.

L'integrale che dobbiamo valutare, ottenuto dopo il cambio di variabile è calcolato lungo il segmento $A'B'$, mentre noi vorremmo ricondurci, come detto, ad un integrale sul segmento AB. Poiché la funzione integranda è olomorfa, a parità di estremi il valore dell'integrale non cambia se si cambia cammino di integrazione (§3.3.2), quindi, per quanto appena detto

$$\frac{1}{2\pi i}\int_{(a-b)-i(h-c)}^{(a-b)+i(h-c)}\frac{x^{t+\beta}}{t}dt$$
$$=\frac{1}{2\pi i}\int_{(a-b)-i(h-c)}^{(a-b)-ih}\frac{x^{t+\beta}}{t}dt+\frac{1}{2\pi i}\int_{(a-b)-ih}^{(a-b)+ih}\frac{x^{t+\beta}}{t}dt$$
$$+\frac{1}{2\pi i}\int_{(a-b)+ih}^{(a-b)+i(h-c)}\frac{x^{t+\beta}}{t}dt.$$

Mostreremo che, per $h\to+\infty$, il contributo del primo e terzo integrale è nullo, in modo che la nostra affermazione risulterà corretta e potremo ricondurci alle formule di Perron.

Consideriamo il primo integrale, per il terzo il ragionamento è analogo: per semplicità poniamo $a-b=k$. Si ha

$$\left|\frac{1}{2\pi i}\int_{k-i(h-c)}^{k-ih}\frac{x^{t+\beta}}{t}dt\right|=\frac{1}{2\pi i}\int_{k-i(h-c)}^{k-ih}\left|\frac{x^{t+\beta}}{t}\right||dt|=\frac{1}{2\pi i}\int_{k-i(h-c)}^{k-ih}\frac{x^k}{|t|}|dt|\le\frac{x^k}{2\pi i}\cdot\frac{c}{\sqrt{k^2+h^2}},$$

che tende a zero per $h\to+\infty$ e, come detto, un ragionamento analogo vale per il terzo integrale.

Tornando alla formula completa, quindi, per quanto riguarda i primi tre integrali, essi sono del tipo

$$\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{x^s}{s-\beta}ds=\frac{1}{2\pi i}\int_{(a-b)-ih}^{(a-b)+ih}\frac{x^{t+\beta}}{t}dt,$$

che, a questo punto, ci consentono di applicare le formule di Perron ($a-b\in\mathbb{R}$, poiché $a\in\mathbb{R}$ e $b=Re(\beta)\in\mathbb{R}$)

$$\frac{1}{2\pi i}\int_{(a-b)-ih}^{(a-b)+ih}\frac{x^{t+\beta}}{t}dt=\frac{1}{2\pi i}\int_{(a-b)-ih}^{(a-b)+ih}\frac{x^{\beta}x^{t}}{t}dt=x^{\beta}\frac{1}{2\pi i}\int_{(a-b)-ih}^{(a-b)+ih}\frac{x^{t}}{t}dt\to x^{\beta},$$

per $h\to+\infty$ e $\alpha>b=Re(\beta)$.

Applichiamo, dunque, quanto detto ai primi tre integrali nella formula.

$$\psi(x)=\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}-\frac{\zeta'(s)}{\zeta(s)}x^s\frac{ds}{s}$$
$$=\lim_{h\to+\infty}\left(\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{s}{s-1}x^s\frac{ds}{s}\right)-\lim_{h\to+\infty}\left(\sum_{\rho}\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{s}{\rho(s-\rho)}x^s\frac{ds}{s}\right)$$
$$+\lim_{h\to+\infty}\left(\sum_{n=1}^{\infty}\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{s}{2n(2n+s)}x^s\frac{ds}{s}\right)-\lim_{h\to+\infty}\left(\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{\zeta'(0)}{\zeta(0)}x^s\frac{ds}{s}\right).$$

Nel primo

$$\lim_{h\to+\infty}\left(\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{s}{s-1}x^s\frac{ds}{s}\right)=\lim_{h\to+\infty}\left(\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{x^s ds}{s-1}\right)=\lim_{h\to+\infty}\left(\frac{1}{2\pi i}\int_{a-1-ih}^{a-1+ih}\frac{x^{t+1}ds}{t}\right)$$
$$=x\lim_{h\to+\infty}\left(\frac{1}{2\pi i}\int_{a-1-ih}^{a-1+ih}\frac{x^t ds}{t}\right)=x.$$

Nel secondo

$$\lim_{h\to+\infty}\left(\sum_{\rho}\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{s}{\rho(s-\rho)}x^s\frac{ds}{s}\right)=\sum_{\rho}\frac{1}{\rho}\left(\lim_{h\to+\infty}\left(\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{x^s ds}{s-\rho}\right)\right)$$
$$=\sum_{\rho}\frac{1}{\rho}\left(\lim_{h\to+\infty}\left(\frac{1}{2\pi i}\int_{a-Re(\rho)-i(h-Im(\rho))}^{a-Re(\rho)+i(h-Im(\rho))}\frac{x^{t+\rho}dt}{t}\right)\right)$$
$$=\sum_{\rho}\frac{x^\rho}{\rho}\left(\lim_{h\to+\infty}\left(\frac{1}{2\pi i}\int_{a-Re(\rho)-ih}^{a-Re(\rho)+ih}\frac{x^t dt}{t}\right)\right)=\sum_{\rho}\frac{x^\rho}{\rho}.$$

Siccome la serie considerata è convergente, abbiamo inteso il limite della somma come la somma dei limiti.

Passiamo al terzo integrale.

$$\lim_{h\to+\infty}\left(\sum_{n=1}^{\infty}\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{s}{2n(2n+s)}x^s\frac{ds}{s}\right)=\sum_{n=1}^{\infty}\frac{1}{2n}\left(\lim_{h\to+\infty}\left(\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{x^s ds}{s+2n}\right)\right)$$
$$=\sum_{n=1}^{\infty}\frac{1}{2n}\left(\lim_{h\to+\infty}\left(\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{x^s ds}{s+2n}\right)\right)$$
$$=\sum_{n=1}^{\infty}\frac{1}{2n}\left(\lim_{h\to+\infty}\left(\frac{1}{2\pi i}\int_{a+2n-ih}^{a+2n+ih}\frac{x^{t-2n}ds}{t}\right)\right)$$
$$=\sum_{n=1}^{\infty}\frac{x^{-2n}}{2n}\left(\lim_{h\to+\infty}\left(\frac{1}{2\pi i}\int_{a+2n-ih}^{a+2n+ih}\frac{x^t ds}{t}\right)\right)=\sum_{n=1}^{\infty}\frac{x^{-2n}}{2n}.$$

Conclusione

Prima di passare alla formula completa, possiamo notare che, per il quarto integrale

$$\lim_{h\to+\infty}\left(\frac{1}{2\pi i}\int_{a-ih}^{a+ih}\frac{\zeta'(0)}{\zeta(0)}x^s\frac{ds}{s}\right)=\lim_{h\to+\infty}\left(\frac{\zeta'(0)}{\zeta(0)}\left(\frac{1}{2\pi i}\int_{a-ih}^{a+ih}x^s\frac{ds}{s}\right)\right)=\frac{\zeta'(0)}{\zeta(0)},$$

per la formula di Perron, in quanto $x>1$.

Possiamo, dunque, concludere con il seguente

Teorema (von Mangoldt)

Per $x>1$ non intero,

$$\psi(x)=x-\sum_{\rho}\frac{x^\rho}{\rho}+\sum_{n=1}^{\infty}\frac{x^{-2n}}{2n}-\frac{\zeta'(0)}{\zeta(0)}.$$

14.3.6 Commenti

Questo risultato è chiamato anche "formula esplicita" (per la ψ) proprio perché è una relazione che ci consente di calcolare esplicitamente i valori della funzione ψ.

Si possono fare varie osservazioni a suo proposito.

(i) I termini dopo la x, possono intendersi come "correttivi" della differenza tra la funzione $\psi(x)$ e la x stessa. Vedremo, nell'Appendice III che questo risultato è alla base della dimostrazione del fatto che

$$\lim_{x\to+\infty}\frac{\psi(x)}{x}=1,$$

risultato alla base della dimostrazione del teorema dei numeri primi.

(ii) Il termine $\zeta'(0)/\zeta(0)$ è ovviamente costante. Riportiamo, per completezza

$$\frac{\zeta'(0)}{\zeta(0)}=\log(2\pi).$$

Rimandiamo chi fosse interessato a conoscerne meglio il valore a ([9], §3.8) o, più brevemente, ancora in ([9], §6.8) (che però sono concettualmente più difficili).

(iii) Oltre a questa formula "esplicita", possiamo segnalare anche la "formula esplicita troncata" (per es. ([28], §6.5)) che si può dimostrare sia a partire da questa sia in maniera indipendente.

$$\psi(x)=x-\sum_{|\rho|\le T}\frac{x^{\rho}}{\rho}+O\left(\frac{x}{T}(\log(xT))^2\right),$$

per $x>1$ (anche intero) e $T>1$. Se ne possono trovare alcune varianti – nelle quali le differenze sono costanti numeriche nell'O grande – che fanno in modo che tale formula sia valida per qualsiasi $x>1$ anche intero.

Concludiamo rimarcando che tale formula "esplicita" per la $\psi(x)$ non costituisce solo l'ennesimo collegamento tra la funzione ζ e i primi, ma mostra un primo collegamento <u>tra gli zeri</u> (non banali) <u>della funzione ζ e i primi</u>.

15. ALTRI RISULTATI PER LA FUNZIONE ζ

La funzione ζ è difficile da valutare, soprattutto lungo la striscia critica, nonostante le sue svariate rappresentazioni. Altrettanto vale per la ξ.
In questa sezione presenteremo alcuni metodi sviluppati per ovviare a questi inconvenienti (senza dimostrazione): si tratta di formule – genericamente – approssimate, che tuttavia spesso consentono ottime stime.

15.1 Formula di Eulero-McLaurin

Ricordiamo la definizione relativa ai numeri di Bernoulli, vista nella sezione dedicata ai prolungamenti analitici della funzione ζ.
Avevamo visto che la funzione

$$f(s) = \frac{s}{e^s - 1}, \qquad s \in \mathbb{C},$$

ammetteva, per $|s| < 2\pi i$, uno sviluppo in serie di Taylor della forma

$$\frac{s}{e^s - 1} = \sum_{n=0}^{\infty} \frac{B_n s^n}{n!}$$

dove i coefficienti B_n sono, per definizione, i numeri di Bernoulli ($n \in \mathbb{N}$). Si era visto, inoltre, che $B_n = 0$ per $n > 1$ dispari.
A partire da essi si costruiscono i così detti polinomi di Bernoulli, nel modo seguente

$$B_n(x) = \sum_{j=0}^{n} \binom{n}{j} B_{n-j} x^j,$$

dove ricordiamo che $\binom{n}{j}$ è il coefficiente binomiale definito nel modo usuale

$$\binom{n}{j} = \frac{n!}{j!\,(n-j)!}.$$

Vediamo, dunque, la valutazione di Eulero-McLaurin della $\zeta(s)$. Per $N > 1$,

$$\begin{aligned}\zeta(s) &= \sum_{n=1}^{N-1} n^{-s} + \int_N^{\infty} x^{-s} dx + \frac{1}{2} N^{-s} - s \int_N^{\infty} B_1(\{x\}) x^{-s-1} dx \\ &= \sum_{n=1}^{N-1} n^{-s} + \frac{N^{1-s}}{s-1} + \frac{1}{2} N^{-s} + \frac{B_2}{2} s N^{-s-1} + \cdots \\ &\quad + \frac{B_{2q}}{(2q)!} s(s+1)\cdots(s+2q-2) N^{-s-2q+1} + R_{2q}.\end{aligned}$$

Si può provare che

$$R_{2q} = -\frac{s(s+1)\cdots(s+2p-1)}{(2q)!} \int_N^{\infty} B_{2q}(\{x\}) x^{-s-2q} dx.$$

Questo resto nasce dal fatto che l'integrale

$$-s\int_N^\infty B_1(\{x\})x^{-s-1}dx$$

non possiede una valutazione "elementare", dunque si itera l'integrazione per parti – fermandosi ad un indice q – per ottenere

$$-s\int_N^\infty B_1(\{x\})x^{-s-1}dx = \sum_{k=1}^{q}\frac{B_{2k}}{(2k)!}\left(f^{(2k-1)}(n) - f^{(2k-1)}(0)\right) + R_{2q},$$

in cui, nel nostro caso, $f(x) = x^{-s}$.

La presenza dei termini pari nella successione dei numeri di Bernoulli è dovuta al fatto che, come detto, quelli dispari sono nulli (a parte il primo che, però, non compare).

La forza di questa serie è che decresce molto rapidamente: si può provare che

$$|R_{2q}| \leq \left|\frac{s(s+1)\cdots(s+2q+1)B_{2(q+1)}N^{-\sigma-2q-1}}{2(q+1)!\,(\sigma+2q+1)}\right|.$$

In essa, infatti, possiamo notare che i termini $N^{-\sigma-2q-1}$ e il fattoriale al denominatore contribuiscono attivamente a questa rapida decrescita.

Si può, dunque, vedere che la formula di Eulero-McLaurin può essere utilizzata per calcolare – con una discreta precisione – i valori della ζ di Riemann. L'ultima espressione, inoltre, fornisce anche la stima dell'errore che si commette in tal senso.

15.2 Equazione funzionale approssimata e formula di Riemann-Siegel

L'equazione funzionale approssimata è un risultato intermedio tra la formula derivata dal metodo di Eulero-McLaurin e la formula di Riemann-Siegel che richiameremo brevemente in seguito.

Siano $x, y \in R^+$, con $2\pi xy = |t|$, allora per $s = \sigma + it$, con $0 \leq \sigma \leq 1$ si ha

$$\zeta(s) = \sum_{n\leq x}\frac{1}{n^s} + \chi(s)\sum_{n\leq y}\frac{1}{n^{1-s}} + O(x^{-\sigma}) + O\left(|t|^{\frac{1}{2}-\sigma}y^{\sigma-1}\right),$$

in cui si è posto

$$\chi(s) = 2^s\pi^{s-1}\sin\left(\frac{\pi s}{2}\right)\Gamma(1-s).$$

Tale equazione funzionale – di cui si può trovare una dimostrazione completa in ([26], §4.1-4.13) (tramite vari lemmi intermedi) – è stata provata da Hardy e Littlewood nel 1922.

A partire da questa, possiamo trovare la seguente formula di Riemann-Siegel (anche se non in forma completa): la dimostrazione completa richiede la valutazione dello stesso integrale visto nella sezione dell'equazione funzionale della zeta lungo la curva chiusa C_N (percorsa in senso antiorario), contenente i punti $\pm 2n\pi i, n = 0,1,\dots,N$. Si ha dunque

$$\zeta\left(\frac{1}{2}+it\right)=\Xi(t)$$
$$=2\sum_{n=1}^{N}\frac{cos(\theta(t)-t\log(n))}{n^{1/2}}$$
$$+\frac{e^{i\theta(t)}e^{-t\pi/2}}{(2\pi)^{1/2+it}e^{-t\pi/4}(1-ie^{t\pi})}\int_{C_N}\frac{(-x)^{-1/2+it}e^{-Nx}dx}{e^{x}-1},$$

nella quale $\theta(t)$ è la funzione θ di Riemann-Siegel ([10]), definita come

$$\theta(t)=\arg\left(\Gamma\left(\frac{2it+1}{4}\right)\right)-\frac{\log(\pi)}{2}t,$$

anche se, generalmente, si utilizza il suo sviluppo asintotico ([26])

$$\theta(t)\sim\frac{t}{2}\log\left(\frac{t}{2\pi}\right)-\frac{t}{2}-\frac{\pi}{8}+\frac{1}{48t}+\frac{t}{5760t^3}+O(t^{-5}).$$

Questa formula era già stata scoperta da Riemann, ma mai pubblicata. Fu Siegel che, a partire dagli scritti personali di Riemann, la provò nel 1932. Una dimostrazione completa si può trovare in ([9], §7.2).

Tale formula fornisce un miglioramento della somma di Eulero-McLaurin per l'approssimazione dei valori di $\zeta(s)$: è tuttora largamente utilizzata (anche se "riadattata") nell'algoritmo di Odlyzko-Schönhage ([31]) per il calcolo degli zeri lungo la linea critica. Non fornisce tuttavia spunti per la dimostrazione dell'ipotesi di Riemann.

16. FORMULA PER LA FUNZIONE π

<<Grazie a questi metodi, il numero dei primi che sono più piccoli di x può essere determinato.>>

(Tratto dall'articolo di Riemann)

Dopo aver discusso riguardo alla zeta e agli zeri della stessa, Riemann passa a quello che è il vero obiettivo del suo articolo di ricerca, cioè la "determinazione del numero dei primi minori di una certa quantità data x" e, dunque, una formula per la $\pi(x)$ che, come vedremo, è strettamente collegata agli zeri (non banali) della ζ.
In questa sezione non dimostreremo a fondo tutto ciò che andrebbe effettivamente dimostrato. Ci saranno, più che altro, dei cenni sulle idee che fondano le dimostrazioni stesse che hanno l'obiettivo di cogliere il ragionamento di Riemann senza perdersi in formalismi eccessivamente complicati che potrebbero anche fuorviare da quelli che sono gli argomenti esposti.
Inizieremo, dunque, parlando della funzione $J(x)$, introdotta dal matematico tedesco (con il nome "comune" $f(x)$) nel suo articolo e strettamente collegata alla $\pi(x)$. Dopo averne visto una formula esplicita, vedremo come, da essa, riusciremo a ottenere una formula concreta per il calcolo della $\pi(x)$.

16.1 Le trasformate di Fourier

Serie e trasformate di Fourier sono argomenti molto importanti nell'ambito dell'analisi matematica: una loro trattazione completa richiederebbe – da sola – una tesi appropriata, dedicata anche alle loro applicazioni, soprattutto in fisica.
Le serie di Fourier, infatti, consentono di scomporre funzioni periodiche complesse in somme di onde semplici mentre le trasformate di Fourier generalizzano le omonime serie trovando spazio praticamente in ogni ambito nel quale compaiono applicazioni complesse dell'analisi matematica.
Ci limiteremo, dunque, ad enunciare i caratteri principali delle trasformate di Fourier.
Consideriamo, inizialmente, il seguente insieme, detto "insieme – o *spazio* – delle funzioni di Schwartz",

$$S(\mathbb{R}) = \{f \in C^{\infty}(\mathbb{R}): |x^n f^{(n)}(n)| < C_n\},$$

dove C_n sono opportune costanti reali. In altre parole le funzioni di Schwartz hanno derivate (di ogni ordine) limitate.
Una funzione in tale insieme – detta, per l'appunto, funzione di Schwartz – è tale che, fissato n intero positivo, $f^{(n)}(x) = O(x^{-n})$. Ovviamente, per $n = 0$ si intende $f^{(0)}(x) = f(x)$ in analogia a quanto accade per la formula di Taylor (§1.3.2).

Si possono dimostrare alcuni risultati interessanti per le funzioni di Schwartz che enunciamo per completezza:

- le funzioni di Schwartz sono limitate;
- le derivate (di ogni ordine) delle funzioni di Schwartz tendono a zero per $x \to \pm\infty$;
- se $f \in S(\mathbb{R})$ e $g \in S(\mathbb{R})$, allora anche $f \pm g, fg, cf \in S(\mathbb{R})$ con c costante reale.

Lo spazio di Schwartz, anche per queste proprietà, è detto anche "spazio delle funzioni a rapida decrescenza". Inoltre, scelta una qualunque funzione $f(x) \in S(\mathbb{R})$, essa è tale che l'integrale

$$\int_{-\infty}^{+\infty} f(x)e^{-i\lambda x}\,dx,$$

è convergente $\forall\lambda \in \mathbb{R}$..

Questo integrale è anche detto "integrale di Fourier" e la funzione

$$\hat{f}(\lambda) = \int_{-\infty}^{+\infty} f(x)e^{-i\lambda x}\,dx$$

è detta "trasformata di Fourier di f" e si indica anche con $\mathcal{F}(f)$.

Enunciamo, infine, la seguente formula di inversione.

Teorema (Formula di Inversione)

Sia $f \in S(\mathbb{R})$, e $\hat{f}$ la sua trasformata di Fourier. Allora $\hat{f} \in S(\mathbb{R})$ e, inoltre, vale la formula di inversione

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{f}(t)e^{ixt}\,dt.$$

Questa formula di inversione ci consente di "tornare indietro" dalla trasformata di Fourier di f alla funzione di partenza: per tale motivo è anche detta "antitrasformata".

16.2 Definizione della $J(x)$

Riportiamo un breve passaggio tratto dall'articolo di Riemann (§Appendice I).

> <<Sia $F(x)$ uguale a questo numero [cioè uguale a $\pi(x)$, *n.d.A.*] quando x non è esattamente uguale ad un numero primo; ma sia più grande di $\frac{1}{2}$ quando x sia un numero primo, allora, per ogni x a cui qui c'è un salto nel valore di $F(x)$ ["per ogni x per il quale c'è un salto del valore di $F(x)$" è da intendersi "per x primo": come vedremo a breve, con la scrittura $F(x)$ Riemann intende (all'incirca) la funzione $\pi(x)$, *n. d. A.*],
>
> $$F(x) = \frac{F(x+0) + F(x-0)}{2}.$$
>
> Se nell'identità
>
> $$\log\zeta(s) = -\sum\log(1-p^s) = \sum p^{-s} + \frac{1}{2}\sum p^{-2s} + \frac{1}{3}\sum p^{-3s} + \cdots$$
>
> adesso si sostituisce

$$p^{-s} \text{ con } s \int_p^\infty x^{-s-1} dx, \quad p^{-2s} \text{ con } s \int_{p^2}^\infty x^{-s-1} dx, \qquad \ldots,$$

uno ottiene

$$\frac{\log \zeta(s)}{s} = \int_1^\infty f(x) x^{-s-1} dx,$$

se uno indica

$$F(x) + \frac{1}{2} F\left(x^{\frac{1}{2}}\right) + \frac{1}{3} F\left(x^{\frac{1}{3}}\right) + \cdots$$

tramite $f(x)$.>>

(Tratto dall'articolo di Riemann)

Queste righe, così come tutto l'articolo nel suo complesso, sono tutt'altro che facili da leggere; tuttavia – al contrario di passaggi decisamente più oscuri e criptici – una giusta chiave di lettura rende l'interpretazione molto semplice.

Riemann inizia con il richiamare una *versione* della funzione $\pi(x)$, nel suo articolo indicata con $F(x)$. Il matematico tedesco, infatti, la intende diversamente dal modo in cui siamo abituati a vederla. Possiamo riassumerla come segue:

$$F(x) = \begin{cases} \pi(x) - \frac{1}{2}, & x \text{ primo} \\ \pi(x), & \text{altrimenti} \end{cases}.$$

Scrittura che Riemann rende ancora più complicata assimilandola ad una media fatta su un intervallo infinitesimo:

$$F(x) = \frac{F(x+0) + F(x-0)}{2}.$$

Tale scrittura, oltre ad essere "originale" – per non dire *sbagliata* – è da intendersi nel modo che segue:

- $F(x) = \pi(x)$ per x diverso da un qualsiasi numero primo;
- $F(x)$ è la media dei due valori assunti in un intervallo infinitesimo $[x - \varepsilon, x + \varepsilon]$ quando x è un numero primo ($\varepsilon > 0$).

Vediamo di fare un esempio.

Scegliendo $x = 10$, abbiamo $F(10) = \pi(10) = 4$: tale valore resta lo stesso anche per $x \in [10,11[$ per definizione della funzione π per i reali. Per $x \in]11,12]$, $F(x) = \pi(x) = 5$.

Tuttavia, per $x = 11$, $\pi(x) = 5$ ma $F(x) = 4{,}5$. Per Riemann, questo si ricava da

$$\lim_{\varepsilon \to 0} \frac{F(11+\varepsilon) + F(11-\varepsilon)}{2} = \frac{4+5}{2} = 4{,}5.$$

In parole povere, la funzione $F(x)$ di Riemann è una $\pi(x)$ pesata che assume un valore intermedio nel "salto" che si avrebbe in prossimità con un numero primo.

Consideriamo, dunque, $s \in \mathbb{C}$ con $Re(s) > 1$. Utilizzando la formula del prodotto di Eulero (§10.1.6) per la funzione $\zeta(s)$, vista in, otteniamo

$$\log(\zeta(s)) = -\log\left(1 - \frac{1}{p^s}\right) = \sum_{p \text{ primo}} \sum_{n=1}^{\infty} \frac{1}{np^{-ns}},$$

dove quest'ultima si ottiene tenendo conto dello sviluppo di Taylor della funzione logaritmo valido per $|s| < 1$,

$$\log(1-s) = -s - \frac{1}{2}s^2 - \frac{1}{3}x^3 - \cdots,$$

direttamente ricavabile da quello usuale per il logaritmo reale, così come si è detto in (§11). Abbiamo, dunque, ottenuto lo stesso risultato di Riemann, anche se il matematico tedesco, nel suo articolo, lo "esplicita" evitando di servirsi della doppia sommatoria:

$$\sum_{p \text{ primo}} \sum_{n=1}^{\infty} \frac{1}{np^{-ns}} = \sum_{p \text{ primo}} \frac{1}{p^s} + \frac{1}{2} \sum_{p \text{ primo}} \frac{1}{p^{2s}} + \frac{1}{3} \sum_{p \text{ primo}} \frac{1}{p^{3s}} + \cdots.$$

Come dice lo stesso Riemann, in seguito si pone

$$p^{-s} = s\int_p^{\infty} x^{-s-1}ds\,, \qquad p^{-2s} = s\int_{p^2}^{\infty} x^{-s-1}ds\,, \qquad \ldots,$$

in quanto non è difficile provare che

$$s\int_{p^n}^{\infty} x^{-s-1}dx = s\left(-\frac{x^{-s}}{s}\bigg|_{p^n}^{\infty}\right) = p^{-ns},$$

per ottenere, successivamente, quanto riportato da Riemann nel suo articolo:

$$\sum_{p \text{ primo}} \sum_{n=1}^{\infty} \frac{1}{np^{-ns}} = \sum_{p \text{ primo}} \sum_{n=1}^{\infty} \frac{s}{n}\int_{p^n}^{\infty} x^{-s-1}dx = s\int_0^{\infty} J(x)x^{-s-1}dx,$$

cioè

$$\frac{\log\big(\zeta(s)\big)}{s} = \int_0^{\infty} J(x)x^{-s-1}dx,$$

in cui con $J(x)$ si indica quella che Riemann intende come $f(x)$, cioè

$$J(x) = \pi(x) + \frac{1}{2}\pi\big(x^{1/2}\big) + \frac{1}{3}\pi\big(x^{1/3}\big) + \cdots = \sum_{n=1}^{\infty} \frac{1}{n}\pi\big(x^{1/n}\big).$$

I seguenti commenti possono essere di qualche utilità.

(i) L'utilizzo della "notazione" $J(x)$ in luogo dell'originale $f(x)$ è posteriore all'articolo di Riemann ed è verosimilmente dovuto al fatto che, in matematica, si utilizza $f(x)$ per indicare una *generica* funzione.

(ii) In realtà, l'ultima uguaglianza nella

$$\sum_{p \text{ primo}} \sum_{n=1}^{\infty} \frac{1}{np^{-ns}} = \sum_{p \text{ primo}} \sum_{n=1}^{\infty} \frac{s}{n}\int_{p^n}^{\infty} x^{-s-1}dx = s\int_0^{\infty} J(x)x^{-s-1}dx,$$

meriterebbe un ulteriore approfondimento. Bisognerebbe dimostrare, infatti, la seguente proprietà, valida per ogni $g(x)$ integrabile:

$$\sum_{p \text{ primo}} \int_{p^n}^{\infty} g(x)dx = \int_0^{\infty} \pi\big(x^{1/n}\big)g(x)dx.$$

L'obiettivo dei prossimi paragrafi è quello di trovare una formula più "semplice" per il calcolo della funzione $J(x)$ per poi seguire il ragionamento di Riemann che la utilizza per la ricerca di una analoga espressione per la $\pi(x)$.

16.3 L'inversione di Fourier

Dopo aver trovato questa relazione tra la $J(x)$ e la $\zeta(s)$, Riemann va oltre.

<<Se, però, l'equazione

$$g(s) = \int_0^\infty h(x)x^{-s} d\log x$$

è valida in mezzo a questo intervallo, allora, facendo uso del teorema di Fourier, si può esprimere la funzione h in termini della funzione g. L'equazione si decompone, se $h(x)$ è reale e

$$g(a+bi) = g_1(b) + ig_2(b),$$

nelle seguenti due:

$$g_1(b) = \int_0^\infty h(x)x^{-a}\cos(b\log x)\, d\log x,$$

$$ig_2(b) = -i\int_0^\infty h(x)x^{-a}\sin(b\log x)\, d\log x.$$

Se uno moltiplica entrambe le equazioni con

$$(\cos(b\log y) + i\sin(b\log y))db$$

e li integra da $-\infty$ a $+\infty$, allora uno ottiene $\pi h(y)y^{-\alpha}$ a destra di entrambe, in accordo dai teoremi di Fourier; così, se si aggiunge entrambe le equazioni e le moltiplica per iy^α, uno ottiene

$$2\pi i h(y) = \int_{a-\infty i}^{a+\infty i} g(s)y^s ds,$$

dove l'integrazione è svolta affinché la parte reale di s resti costante.
Per un valore di y in cui c'è un salto nel valore di $h(y)$, l'integrale porta alla media dei valori della funzione h ad ogni lato del salto. Dal modo in cui la funzione f è definita, vediamo che essa ha la stessa proprietà, e, in generale,

$$f(y) = \frac{1}{2\pi i}\int_{a-\infty i}^{a+\infty i} \frac{\log\zeta(s)}{s} y^s\, ds.>>$$

(Tratto dall'articolo di Riemann)

Dopo essere stato piuttosto sintetico in altri punti del suo articolo di ricerca, stavolta Riemann ha cura di "richiamare" le trasformate di Fourier che utilizzerà subito dopo per esplicitare la funzione $J(x)$ portandola fuori dall'integrale.
Avevamo, infatti, trovato la formula

$$\frac{\log(\zeta(s))}{s} = \int_0^\infty J(x)x^{-s-1}dx,$$

che Riemann indica con

$$\frac{\log(\zeta(s))}{s} = \int_0^\infty J(x)x^{-s} d\log(x),$$

scrittura che è analoga alla precedente in quanto, nell'integrale di Stieltjes (§Appendice III), se la funzione integratrice $g(x)$ è almeno C^1, si ha

$$dg(x) = g'(x)dx,$$

nel nostro caso

$$d\log(x) = x^{-1}dx.$$

Ora si tratta di riscrivere la funzione $J(x)$ in termini della $\zeta(s)$ applicando l'inversione di Fourier richiamata da Riemann nel suo articolo e da noi ad inizio sezione. Il matematico tedesco, infatti, inizia dal porre $s = a + ib$, nel quale $a > 1$ è una costante reale e $\lambda = \log(x)$, da cui $x = e^{\lambda}$. La formula ottenuta in precedenza, dunque, diventa

$$\frac{\log\big(\zeta(a+ib)\big)}{a+ib} = \int_{-\infty}^{+\infty} J\big(e^{\lambda}\big)e^{-(a+ib)\lambda}d\lambda\,, = \frac{1}{2\pi}\int_{-\infty}^{+\infty} 2\pi J\big(e^{\lambda}\big)e^{-(a+ib)\lambda}d\lambda$$

tenendo conto del fatto che anche gli estremi cambiano al cambio di variabile (se λ variava da 0 a $+\infty$, $x = e^{\lambda}$ varia da $-\infty$ a $+\infty$), mentre nell'ultimo passaggio si è moltiplicato e diviso per 2π.

A questo punto possiamo applicare la formula di inversione alla funzione $2\pi J(e^{x})e^{-ax}$ (nell'integrale la variabile di integrazione è λ), ottenendo

$$2\pi J(e^{x})e^{-ax} = \int_{-\infty}^{+\infty} \frac{\log\big(\zeta(a+ib)\big)}{a+ib} e^{bx}\,db,$$

da cui, cambiando nuovamente variabile – cioè ponendo $y = e^{x}$ – si ottiene

$$J(y) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{\log\big(\zeta(a+ib)\big)}{a+ib} y^{a+ib}\,db.$$

Dal fatto che a è costante, possiamo concludere

$$J(x) = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} \log\big(\zeta(s)\big)\,x^{s}\frac{ds}{s},$$

dove si è posto $y = x$ per questioni puramente estetiche (Riemann, nel suo articolo, utilizza ancora l'incognita y).

16.4 Sostituzione nell'integrale

L'obiettivo è ora quello di trovare una forma esplicita per la funzione $J(x)$. L'ultima rappresentazione, cioè

$$J(x) = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} \log\big(\zeta(s)\big)\,x^{s}\frac{ds}{s},$$

si serve di un'integrale che occorre valutare meglio (servendosi della formula di Perron come ci suggerisce la forma stessa dell'integrale nel suo complesso).

Il ragionamento è analogo a quello utilizzato per la determinazione della formula esplicita per la ψ (§14.3) e lo si può ritrovare, ampiamente spiegato, anche nell'articolo di Riemann. Il matematico tedesco, infatti, sorvola su questioni "secondarie" come la formula prodotto per la ξ o la disposizione degli zeri della stessa per focalizzarsi su ciò che gli importa davvero: la "determinazione del numero dei primi inferiori ad una quantità data x".

Si tratta, dunque, di prendere da un lato la definizione stessa della ξ derivabile dall'equazione funzionale della ζ (§13.1.1) utilizzando la funzione $\prod(s)$ invece della $\Gamma(s)$ (per semplificare i calcoli successivamente)

$$\xi(s) = \prod\left(\frac{s}{2}\right)\pi^{-\frac{s}{2}}(s-1)\zeta(s),$$

e, dall'altro, la formula prodotto per la ξ (§Appendice III)

$$\xi(s) = \xi(0)\prod_{\rho}\left(1-\frac{s}{\rho}\right),$$

nel quale ρ indica gli zeri della ξ (cioè gli zeri non banali della ζ). Passando al logaritmo ambo i membri (fissando un ramo regolare dello stesso) e isolando $\log(\zeta(s))$, otteniamo

$$\log(\zeta(s)) = \log(\xi(0)) + \sum_{\rho}\log\left(1-\frac{s}{\rho}\right) - \log\left(\prod\left(\frac{s}{2}\right)\right) + \frac{s}{2}\log(\pi) - \log(s-1).$$

Tale sostituzione sarà rimandata al prossimo paragrafo poiché, stavolta, una tale scrittura crea subito problemi: sostituendola direttamente nella

$$J(x) = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\log(\zeta(s))\,x^s\frac{ds}{s},$$

otteniamo integrali divergenti.

Riemann lo sapeva, per questo fa un ulteriore passo in avanti, integrando per parti. A partire dalla

$$J(x) = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\log(\zeta(s))\,x^s\frac{ds}{s},$$

sostituendo $x^s = e^{s\log(x)}$, si può notare che

$$\frac{d}{ds}(x^s) = \frac{d}{ds}\left(e^{s\log(x)}\right) = \log(x)\,e^{s\log(x)} = \frac{x^s}{\log(x)},$$

da cui

$$J(x) = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\log(\zeta(s))\,x^s\frac{ds}{s} = \frac{1}{2\pi i\log(x)}\int_{a-i\infty}^{a+i\infty}\log(\zeta(s))\frac{d}{ds}(x^s)\frac{ds}{s},$$

che si può integrare agevolmente per parti secondo la regola usuale

$$\int_a^b f'(s)g(s)ds = f(s)g(s)|_a^b - \int_a^b f(s)g'(s)ds,$$

nella quale

$$f(s) = x^s, \qquad g(s) = \frac{\log(\zeta(s))}{s}, \qquad \left(f'(s) = \frac{d}{ds}(x^s)\right).$$

Abbiamo, dunque, con $a > 1$

$$J(x) = \frac{1}{2\pi i\log(x)}\int_{a-i\infty}^{a+i\infty}\log(\zeta(s))\frac{d}{ds}(x^s)\frac{ds}{s}$$
$$= \frac{1}{2\pi i\log(x)}\frac{\log(\zeta(s))}{s}x^s\Big|_{a-i\infty}^{a+i\infty} - \frac{1}{2\pi i\log(x)}\int_{a-i\infty}^{a+i\infty}x^s\frac{d}{ds}\left(\frac{\log(\zeta(s))}{s}\right)ds.$$

Tale integrale diventa

$$J(x) = -\frac{1}{2\pi i\log(x)}\int_{a-i\infty}^{a+i\infty}x^s\frac{d}{ds}\left(\frac{\log(\zeta(s))}{s}\right)ds,$$

poiché

$$\lim_{T\to\pm\infty}\left(\frac{\log\big(\zeta(a\pm iT)\big)}{a\pm iT}x^{a\pm iT}\right)=0.$$

Proviamo brevemente quest'ultimo risultato (prima di procedere).
Possiamo notare che, per $x>1$,

$$\big|\log\big(\zeta(a\pm iT)\big)\big|=\left|\sum_{p\text{ primo}}\sum_{n=1}^{\infty}\frac{1}{np^{-ns}}\right|\leq\sum_{p\text{ primo}}\sum_{n=1}^{\infty}\left|\frac{1}{np^{-ns}}\right|=\sum_{p\text{ primo}}\sum_{n=1}^{\infty}\frac{1}{np^{-na}}$$
$$=\log\big(\zeta(a)\big),$$

in cui l'ultimo termine è una costante e non dipende dalla variabile portata al limite (cioè T, la parte immaginaria di s).
Per quanto riguarda $(a\pm iT)$, notiamo facilmente che tale termine tende – in modulo – a $+\infty$ per $T\to+\infty$ e, trovandosi al denominatore, rende infinitesima la frazione. Infine,

$$\left|x^{a\pm iT}\right|=|x^a|\cdot\left|x^{\pm iT}\right|=x^a,$$

che, anche in questo caso, è una costante rispetto alla variabile T. Dunque, complessivamente, si ha

$$\lim_{T\to\pm\infty}\left(\frac{\log\big(\zeta(a\pm iT)\big)}{a\pm iT}x^{a\pm iT}\right)=0.$$

A questo punto, otteniamo

$$J(x)=-\frac{1}{2\pi i\log(x)}\int_{a-i\infty}^{a+i\infty}x^s\frac{d}{ds}\left(\frac{\log\big(\zeta(s)\big)}{s}\right)ds,$$

cioè la stessa formula di Riemann nel suo articolo (anche se con l'intera derivata al numeratore) (§Appendice I).

16.5 Formula per la $J(x)$

Riemann giunge al seguente risultato

$$J(x)=Li(x)-\sum_{Im(\rho)>0}[Li(x^{\rho})+Li(x^{1-\rho})]+\int_x^{\infty}\frac{dt}{t(t^2-1)\log(t)}+\log\big(\xi(0)\big),\qquad x>1,$$

che sarà il punto di partenza per arrivare alla formula definitiva per la funzione $\pi(x)$.
Come detto nel paragrafo precedente, il procedimento è concettualmente simile a quello che conduce alla determinazione della formula esplicita della ψ: ci limiteremo a vederlo nei suoi passi fondamentali. A partire da

$$J(x)=-\frac{1}{2\pi i\log(x)}\int_{a-i\infty}^{a+i\infty}x^s\frac{d}{ds}\left(\frac{\log\big(\zeta(s)\big)}{s}\right)ds,$$

si sostituisce all'integrando la seguente

$$\log\big(\zeta(s)\big)=\log\big(\xi(0)\big)+\sum_{\rho}\log\left(1-\frac{s}{\rho}\right)-\log\left(\prod\left(\frac{s}{2}\right)\right)+\frac{s}{2}\log(\pi)-\log(s-1),$$

ricavata dall'eguagliare le due formulazioni per la ξ (quella dall'equazione funzionale della ζ e quella della formula prodotto) isolando il $\log\big(\zeta(s)\big)$. Il risultato è il seguente

$$
\begin{aligned}
J(x) = -\frac{1}{2\pi i \log(x)} \int_{a-i\infty}^{a+i\infty} x^s \frac{d}{ds}\left(\frac{\log\big(\zeta(s)\big)}{s}\right) ds \\
= -\frac{1}{2\pi i \log(x)} \int_{a-i\infty}^{a+i\infty} x^s \frac{d}{ds}\left(\frac{\log\big(\xi(0)\big)}{s}\right) ds \\
- \frac{1}{2\pi i \log(x)} \int_{a-i\infty}^{a+i\infty} x^s \frac{d}{ds}\left(\frac{\sum_\rho (\log(1-s/\rho))}{s}\right) ds \\
+ \frac{1}{2\pi i \log(x)} \int_{a-i\infty}^{a+i\infty} x^s \frac{d}{ds}\left(\frac{\prod(s/2)}{s}\right) ds \\
- \frac{1}{2\pi i \log(x)} \int_{a-i\infty}^{a+i\infty} x^s \frac{d}{ds}\left(\frac{1}{2}\log(\pi)\right) ds \\
+ \frac{1}{2\pi i \log(x)} \int_{a-i\infty}^{a+i\infty} x^s \frac{d}{ds}\left(\frac{\log(s-1)}{s}\right) ds.
\end{aligned}
$$

I calcoli sono molto lunghi e lo stesso Riemann, nel suo articolo, si limita a dei cenni sul loro svolgimento evidenziandone i punti fondamentali: come detto, anche qui riporteremo solo i risultati, analizzando solo alcuni aspetti.

Per chi fosse interessato, rimandiamo alla dimostrazione completa presente in ([9], §1.14-1.17). Innanzitutto, uno dei termini è molto semplice da valutare:

$$
\begin{aligned}
-\frac{1}{2\pi i \log(x)} \int_{a-i\infty}^{a+i\infty} x^s \frac{d}{ds}\left(\frac{\log\big(\xi(0)\big)}{s}\right) ds \\
= -\frac{1}{2\pi i \log(x)} \frac{\log\big(\xi(0)\big)}{s} x^s \Big|_{a-i\infty}^{a+i\infty} + \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} \frac{\log\big(\xi(0)\big)}{s} x^s ds = \log\big(\xi(0)\big),
\end{aligned}
$$

risultato che si ottiene grazie alla formula di Perron ($x > 1$) vista in (§14.2) e dal fatto che

$$
\frac{\log\big(\xi(0)\big)}{s} x^s \Big|_{a-i\infty}^{a+i\infty} = 0,
$$

che si può dedurre con un ragionamento analogo a quanto visto per

$$
\lim_{T\to\pm\infty} \left(\frac{\log\big(\zeta(a \pm iT)\big)}{a \pm iT} x^{a\pm iT}\right) = 0,
$$

sempre ponendo $s = a \pm iT$ (anche se in questo caso, di per sé, $\xi(0) = 1/2$ è già – a prescindere da sostituzioni – una costante).

A questo punto, Riemann si accorge che tutti gli altri integrali sono del tipo

$$
\pm \frac{1}{2\pi i \log(x)} \int_{a-i\infty}^{a+i\infty} x^s \frac{d}{ds}\left(\frac{\log(1-s/\beta)}{s}\right) ds,
$$

per β opportuna costante. L'unico dubbio potrebbe giungere dal

$$
\frac{1}{2\pi i \log(x)} \int_{a-i\infty}^{a+i\infty} x^s \frac{d}{ds}\left(\frac{\prod(s/2)}{s}\right) ds,
$$

ma, nell'integrando, dalla formula prodotto per la $\prod$ (§8.4) si ha

$$
\begin{aligned}
\frac{d}{ds}\left(\frac{\prod(s/2)}{s}\right) = \sum_{n=1}^{\infty} \frac{d}{ds}\left(\frac{-\log(1+s/2n) + \frac{s}{2}\log(1+1/n)}{s}\right) \\
= \sum_{n=1}^{\infty} \frac{d}{ds}\left(\frac{-\log(1+s/2n)}{s} + \frac{1}{2}\log\left(1+\frac{1}{n}\right)\right) = \sum_{n=1}^{\infty} \frac{d}{ds}\left(\frac{-\log(1+s/2n)}{s}\right),
\end{aligned}
$$

in quanto l'altro termine di cui si compone ogni addendo della sommatoria è costante e la sua derivata è dunque nulla.
Anche questo, dunque, è un integrale del tipo

$$\pm\frac{1}{2\pi i\log(x)}\int_{a-i\infty}^{a+i\infty}x^s\frac{d}{ds}\left(\frac{\log(1-s/\beta)}{s}\right)ds.$$

Riemann, quindi, sposta la sua attenzione nello studio di tale integrale: come abbiamo detto in precedenza, il calcolo completo è in ([9], §1.13-1.16).
Si tratta di applicare le proprietà già citate degli integrali in un modo molto simile a quanto detto per provare l'equazione funzionale della ζ o la formula esplicita per la ψ.
Si analizzano dunque i singoli termini per giungere alle seguenti conclusioni.

(i) Il primo termine, detto anche "termine principale" è

$$\frac{1}{2\pi i\log(x)}\int_{a-i\infty}^{a+i\infty}x^s\frac{d}{ds}\left(\frac{\log(s-1)}{s}\right)ds=Li(x).$$

Per giungere a questa formulazione, si dimostra prima di tutto che l'integrando è limitato, per poi applicare la formula di Perron. Tuttavia si passa anche lungo delle funzioni ausiliarie e cambi di variabile servendosi di integrali su curve, in analogia a quanto visto per l'equazione funzionale.

(ii) Il secondo termine è quello riguardante le radici

$$-\frac{1}{2\pi i\log(x)}\int_{a-i\infty}^{a+i\infty}x^s\frac{d}{ds}\left(\frac{\sum_\rho(\log(1-s/\rho))}{s}\right)ds.$$

In questo caso si scompone la somma contando le radici ρ e $1-\rho$ in analogia a quanto visto per la formula prodotto della ξ (§Appendice IV) (nella quale si dimostra anche che tale serie è convergente):

$$\sum_\rho\log\left(1-\frac{s}{\rho}\right)=\sum_{Im(\rho)>0}\left[\log\left(1-\frac{s}{\rho}\right)+\log\left(1-\frac{s}{1-\rho}\right)\right].$$

Si ottengono due integrali da valutare su altrettante curve per poi dedurre, tramite il teorema dei residui

$$-\frac{1}{2\pi i\log(x)}\int_{a-i\infty}^{a+i\infty}x^s\frac{d}{ds}\left(\frac{\sum_\rho(\log(1-s/\rho))}{s}\right)ds$$
$$=-\sum_{Im(\rho)>0}[Li(x^\rho)+Li(x^{1-\rho})].$$

Gli altri termini vengono, generalmente, considerati insieme deducendone il risultato dallo studio del generico integrale

$$\pm\frac{1}{2\pi i\log(x)}\int_{a-i\infty}^{a+i\infty}x^s\frac{d}{ds}\left(\frac{\log(1-s/\beta)}{s}\right)ds.$$

La formula finale che si ottiene è, dunque

$$J(x)=Li(x)-\sum_{Im(\rho)>0}[Li(x^\rho)+Li(x^{1-\rho})]+\int_x^\infty\frac{dt}{t(t^2-1)\log(t)}+\log\big(\xi(0)\big),\qquad x>1.$$

In omaggio a Riemann, lasceremo il termine

$$\log\big(\xi(0)\big),$$

senza necessariamente sostituirlo con

$$\log(\xi(0)) = \log\left(\frac{1}{2}\right) = -\log(2).$$

Ricordiamo, inoltre, che $\xi(0)$ è proprio $\xi(0)$, poiché la sostituzione operata da Riemann trattando la formula prodotto (cioè $s = 1/2 + it$, con $t \in \mathbb{C}$), vale solo in quel frangente (§13.1.6).

16.6 Dalla $J(x)$ alla $\pi(x)$

Dalla

$$J(x) = \sum_{n=1}^{\infty} \frac{1}{n} \pi\left(x^{1/n}\right),$$

ricaviamo

$$\pi(x) = \sum_{n=1}^{\infty} \frac{\mu(n)}{n} J\left(x^{1/n}\right),$$

ottenuta applicando la generalizzazione della formula di inversione di Möbius (come dice lo stesso Riemann nel suo articolo di ricerca (§Appendice I)), di cui si è parlato in (§10.1.3).
Possiamo notare, inoltre, che per x fissato la serie è finita: infatti esiste n_0 intero positivo tale che, per ogni $n > n_0$ intero

$$x^{\frac{1}{n_0}} \geq 2, \qquad x^{\frac{1}{n}} < 2,$$

che, prendendo la definizione originale, si traduce in

$$J(x) = \sum_{n=1}^{\infty} \frac{1}{n} \pi\left(x^{1/n}\right) = \sum_{n=1}^{n_0} \frac{1}{n} \pi\left(x^{1/n}\right) + \sum_{n=n_0+1}^{\infty} \frac{1}{n} \pi\left(x^{1/n}\right) = \sum_{n=1}^{n_0} \frac{1}{n} \pi\left(x^{1/n}\right),$$

in quanto

$$\pi\left(x^{1/n}\right) = 0, \qquad (\pi(x) = 0, x < 2).$$

Riemann ha, dunque, ottenuto una formula *esatta* per la funzione $\pi(x)$. Tuttavia, senza smorzare l'entusiasmo per un risultato così ambito e rilevante, occorre fare le giuste considerazioni.

(i) Analizziamo, innanzitutto, la formula

$$\pi(x) = \sum_{n=1}^{\infty} \frac{\mu(n)}{n} J\left(x^{1/n}\right),$$

per vederne la "praticabilità".
La funzione $\mu(n)$, per come è definita (§10.1.1), non è semplice da determinare poiché un suo calcolo richiederebbe la fattorizzazione di n. Tuttavia, in questa sommatoria, rispetto all'input x, ci si ferma ad un n molto piccolo (dell'ordine di $\log(x)$) e questo ne riduce notevolmente la complessità.
Resta il calcolo della funzione $J(x)$ a destare qualche dubbio. Tuttavia, per x molto grandi si può eseguire tramite la formula esplicita mentre per x inferiori si

può partire anche dalla definizione diretta avendo a disposizione tavole dei valori di $\pi(x)$.

Per x grandi il calcolo di $\pi(x)$ è molto complicato ed è, dunque, preferibile passare attraverso la formula esplicita anche se questa, di per sé, richiede di calcolare molti integrali. Tuttavia l'analisi numerica ci fornisce numerose formule (di quadratura) che consentono di ottenere con ragionevole rapidità risultati con una buona approssimazione.

(ii) Vedendo la formula esplicita per la $J(x)$, cioè

$$J(x) = Li(x) - \sum_{Im(\rho)>0} [Li(x^{\rho}) + Li(x^{1-\rho})] + \int_x^{\infty} \frac{dt}{t(t^2-1)\log(t)} + \log(\xi(0)),$$

(con $x > 1$), ci si può chiedere come influiscono i vari termini.

$Li(x)$, come detto, è il termine principale in quanto il maggior responsabile dell'andamento della $J(x)$ (vedremo qualche esempio nel prossimo paragrafo). Gli altri termini, invece, sono soltanto dei "correttivi" che servono per "limare" l'errore che si ottiene considerando solo il primo.

(iii) La sommatoria del logaritmo integrale fatta sugli zeri (non banali) della ζ è uno dei termini secondari nella formula della $J(x)$

$$\sum_{Im(\rho)>0} [Li(x^{\rho}) + Li(x^{1-\rho})].$$

E', come appena detto, uno dei correttivi presenti dentro tale formula tuttavia non è affatto limitata. Lehmer, infatti, ha dimostrato che tale serie converge solo in valore assoluto mentre, generalmente, diverge.

(iv) Il termine $\log(\xi(0))$ è l'unica costante in tutta la formula:

$$\log(\xi(0)) = \log\left(\frac{1}{2}\right) = -0{,}69314718 \ldots$$

(v) Si può dimostrare che l'integrale

$$\int_x^{\infty} \frac{dt}{t(t^2-1)\log(t)},$$

è infinitesimo per $x \to +\infty$ e ha un contributo davvero irrilevante per x sufficientemente grande.

In generale, infatti, si considerano numeri primi molto grandi, da cui almeno $x > 10^{10}$. Per l'integrando, ciò vuol dire (siccome $\log(x) > 1$)

$$\int_x^{\infty} \frac{dt}{t(t^2-1)\log(t)} \leq \int_x^{\infty} \frac{dt}{t(t^2-1)} = -\frac{1}{2}\log\left(1 - \frac{1}{x^2}\right),$$

da cui, per $x > 10^{10}$ si conclude

$$-\frac{1}{2}\log\left(1 - \frac{1}{x^2}\right) \leq -\frac{1}{2}\log\left(1 - \frac{1}{10^{20}}\right) \cong 5 \cdot 10^{-21},$$

tuttavia tale contributo è già trascurabile per valori molto inferiori (per $x = 10^2$ il contributo è inferiore a $5 \cdot 10^{-5}$).

16.7 Formula approssimata

Vediamo brevemente una formula approssimata della $\pi(x)$, ricavabile direttamente da quella esplicita trovata in precedenza.
Ripartiamo dalla formula per la $\pi(x)$ trovata da Riemann tramite l'inversione di Möbius a quella della $J(x)$:

$$\pi(x) = \sum_{n=1}^{\infty} \frac{\mu(n)}{n} J\left(x^{1/n}\right),$$

in cui, per la $J(x)$, vale la formula esplicita

$$J(x) = Li(x) - \sum_{Im(\rho)>0} [Li(x^{\rho}) + Li(x^{1-\rho})] + \int_x^{\infty} \frac{dt}{t(t^2-1)\log(t)} + \log\left(\xi(0)\right).$$

Per le osservazioni precedenti, si può concludere

$$J(x) = Li(x) - \sum_{Im(\rho)>0} [Li(x^{\rho}) + Li(x^{1-\rho})] + termine\ trascurabile\ + \log\left(\xi(0)\right),$$

cioè (ipotizzando x sufficientemente grande)

$$J(x) \cong Li(x) - \sum_{Im(\rho)>0} [Li(x^{\rho}) + Li(x^{1-\rho})] + \log\left(\xi(0)\right).$$

La sommatoria, come detto, non è convergente ma è comunque trascurabile rispetto a $Li(x)$. Riemann, dunque, pone

$$J(x) \cong Li(x),$$

per poi concludere

$$\pi(x) = \sum_{n=1}^{\infty} \frac{\mu(n)}{n} J\left(x^{1/n}\right) \cong \sum_{n=1}^{\infty} \frac{\mu(n)}{n} Li\left(x^{1/n}\right).$$

Tale approssimazione innanzitutto giustifica quella ipotizzata da Gauss (§9.2)

$$\pi(x) \cong Li(x),$$

che si ottiene prendendo solo il primo termine di quella sommatoria. Tuttavia, da sola, è decisamente più precisa di quella appena riportata: si consideri, come esempio, la seguente tabella nella quale poniamo

$$\sum_{n=1}^{\infty} \frac{\mu(n)}{n} Li\left(x^{1/n}\right) = r(n),$$

che, come detto, è una somma finita.
Possiamo notare come la stima di Riemann sia migliore rispetto a quella di Gauss; si intravede, inoltre, anche l'aumento (in valore assoluto) dell'errore dovuto a tale stima al crescere di n.

n	$\pi(n)$	$Li(n)$	$\pi(x) - Li(x)$	$r(n)$	$\pi(n) - r(n)$
100	25	30	5	26	1
1000	168	178	10	168	0
10000	1229	1246	17	1227	-2
10^6	9592	9630	38	9587	-5
10^7	78498	78638	130	78527	29
10^8	664579	664918	339	664667	88
10^9	5761455	5762209	754	5761552	97
10^{10}	50847534	50849235	1701	50847455	-79
10^{11}	455052511	455055615	3104	455050683	-1828
10^{12}	4118054813	4118066401	11588	4118052495	-2318

Tuttavia, al contrario della precedente, tale stima oscilla intorno al valore vero, poiché, proprio per come è scritta:

$$r(n) = \sum_{n=1}^{\infty} \frac{\mu(n)}{n} Li\left(x^{1/n}\right),$$

possiede termini positivi e negativi (a causa della $\mu(n)$).

16.8 Importanza di questo risultato

Certamente il risultato ottenuto da Riemann, anche ad una prima analisi, è tutt'altro che di poco conto. Tuttavia, esso acquista decisamente ulteriore importanza se ci si sofferma ad osservarlo più attentamente.

Ripartiamo dalla formula per la $\pi(x)$:

$$\pi(x) = \sum_{n=1}^{\infty} \frac{\mu(n)}{n} J\left(x^{1/n}\right),$$

nel quale

$$J(x) = Li(x) - \sum_{Im(\rho)>0} [Li(x^{\rho}) + Li(x^{1-\rho})] + \int_x^{\infty} \frac{dt}{t(t^2-1)\log(t)} + \log\left(\xi(0)\right).$$

L'importanza sta proprio nella formula della $J(x)$.

Il termine sotto il segno di integrale, come detto, è trascurabile fin da valori piuttosto bassi della x, tuttavia è proprio la sommatoria

$$\sum_{Im(\rho)>0} [Li(x^{\rho}) + Li(x^{1-\rho})],$$

a destare il maggiore interesse.

Come detto, tale formula conferma la precedente stima di Gauss per mezzo del logaritmo integrale, tuttavia, possiamo concludere che gli zeri (non banali) della funzione ζ sono utilizzati come "correttivi" di tale stima.

In altre parole questo è proprio il collegamento più *forte* esistente tra la ζ e i primi: nella formula esplicita della $\pi(x)$ si trova proprio che i termini secondari variano a seconda degli zeri della ζ.

17. CONSEGUENZE DELL'IPOTESI DI RIEMANN

In quest'ultima sezione enunceremo (senza dimostrazione) alcune tra le conseguenze più importanti dell'ipotesi di Riemann.

17.1 Ipotesi di Lindelöf

L'ipotesi di Lindelöf dice che

$$\zeta\left(\frac{1}{2}+it\right)=O(t^{\varepsilon}),$$

per ogni $\varepsilon>0$. La si può estendere nel seguente modo, cioè che

$$\zeta(\sigma+it)=O(t^{\varepsilon}),$$

per ogni $\varepsilon>0$ e $\sigma\geq 1/2$. Se ora definiamo

$$\mu(\sigma)=\inf\{\varepsilon\in\mathbb{R}:\zeta(\sigma+it)=O(t^{\varepsilon})\},$$

l'ipotesi di Lindelöf è equivalente a considerare che

$$\mu(\sigma)=0.$$

Ecco qualche ulteriore chiarimento al riguardo.
Innanzitutto, dire

$$\zeta(\sigma+it)=O(t^{\varepsilon}),$$

equivale ad affermare che la funzione ζ è dello stesso ordine di t^{ε}, per $\varepsilon>0$ fissato. Nella sezione dedicata ai teoremi di von Mangoldt si era dimostrato che

$$|\zeta(s)|=O(\log(t)),$$

solo che tale stima valeva solo per $Re(s)>1$. In quel caso, infatti, la questione è più semplice mentre ora una stima globale risulta complicata a causa del comportamento della ζ nella striscia critica (che è, come detto, la regione "interessante").
Indicando con $N(\sigma,T)$ il numero degli zeri della funzione ζ nella retta $Re(s)=\sigma$, al variare di $T=Im(s)$ lungo la direzione puramente immaginaria (numero che non può essere negativo), possiamo enunciare il seguente

Teorema ([26], §13.5)
Condizione necessaria e sufficiente per la validità dell'ipotesi di Lindelöf è che, per ogni $\sigma>1/2$,

$$N(\sigma,T+1)-N(\sigma,T)=o(\log(T)).$$

Di questo teorema non daremo la dimostrazione. Notiamo tuttavia che esso implica l'equivalenza tra l'ipotesi di Riemann e quella di Lindelöf.

La prima implicazione è semplice da vedere: infatti, se è vera l'ipotesi di Riemann, allora tutti gli zeri non banali della ζ si hanno per $Re(s) = \sigma = 1/2$ e, dunque,

$$N(\sigma, T+1) - N(\sigma, T) = 0 = o(\log(T^0)),$$

in quanto, in essa, $\sigma > 1/2$ (dunque $\sigma \neq 1/2$).
Il viceversa utilizza risultati avanzati di analisi complessa e venne provato da Littlewood nel 1912.
In virtù dell'equivalenza tra le due ipotesi, molte menti matematiche si sono profuse nel dimostrare – in modo indipendente da quella di Riemann – l'ipotesi di Lindelöf.
Tuttavia, fino ad ora, anche questa è priva di una dimostrazione: la stima migliore è quella di Huxley (2004) ([4])

$$\varepsilon = \frac{32}{205} \cong 0{,}1561.$$

17.2 Relazioni con il Teorema dei Numeri Primi

Ricordiamo la definizione della funzione λ di Liouville vista nella sezione di Teoria Analitica dei Numeri. Avevamo definito $\lambda(1) = 1$ mentre, per $n > 1$, scomposto nel prodotto di potenze di fattori primi distinti come

$$n = \prod_{i=1}^{r} p_i^{a_i}, \qquad p_i \text{ primo}, \qquad a_i \geq 1,$$

si poneva

$$\lambda(n) = (-1)^{a_1 + a_2 + \cdots + a_n}.$$

Nel 1899, Landau provò nella sua tesi di dottorato il seguente

Teorema
L'ipotesi di Riemann è equivalente all'affermazione che, per ogni $\varepsilon > 0$ fissato,

$$\lim_{n \to +\infty} \frac{\lambda(1) + \cdots + \lambda(n)}{n^{\frac{1}{2}+\varepsilon}} = 0.$$

Nell'appendice (§Appendice III), dimostreremo che

$$\lim_{x \to +\infty} \frac{\pi(x)}{x/\log(x)} = 1,$$

risultato che prende il nome di Teorema dei Numeri Primi dimostrato per la prima volta da Ch. De la Vallée-Poussin e J. Hadamard. Parte della già citata tesi di dottorando di Landau era dedicata al seguente

Teorema
Il Teorema dei Numeri Primi è equivalente alla seguente affermazione

$$\lim_{n \to +\infty} \frac{\lambda(1) + \cdots + \lambda(n)}{n} = 0.$$

Dal confronto degli ultimi due risultati si vede, tramite la funzione λ di Liouville, che il Teorema dei Numeri Primi è esso stesso una conseguenza dell'ipotesi di Riemann e in questo senso è più debole: corrisponde infatti al caso particolare $\varepsilon = 1/2$ del teorema precedente.
Ricordando, ora, la definizione di logaritmo integrale

$$Li(x) = \int_2^{\infty} \frac{dt}{\log(t)},$$

si può dimostrare che

Teorema
L'ipotesi di Riemann è equivalente a dire che

$$\pi(x) = Li(x) + O\left(\sqrt{x}\log(x)\right).$$

La validità dell'ipotesi di Riemann fornisce, dunque, un termine d'errore più preciso rispetto alle stime viste nella sezione dedicata al logaritmo integrale (§9.2). Inoltre, la stima appena enunciata ci conferma che, nella formula esplicita per la funzione $\pi(x)$, il termine principale è proprio quello dovuto al logaritmo integrale e che gli altri sono dei correttivi trascurabili (anche se divergenti).

17.3 La funzione μ di Möbius

Ricordiamo la definizione della funzione μ di Möbius vista nella sezione di TADN (§10.1.1).
Si poneva, anzitutto

$$\mu(1) = 1.$$

Se poi $n > 1$ si decompone nel prodotto di potenze di fattori primi distinti come

$$n = \prod_{i=1}^{r} p_i^{a_i} = p_1^{a_1} \cdot p_2^{a_2} \cdot \ldots \cdot p_r^{a_r}, \qquad p_i \text{ primo}, \qquad a_i \geq 1,$$

allora

$$\mu(n) = \begin{cases} (-1)^r, & \text{se } a_1 = a_2 = \cdots = a_r = 1, \\ 0, & \text{altrimenti}. \end{cases}$$

A partire dalla μ, definiamo la seguente funzione, detta anche funzione di Mertens: per $x > 1$ reale

$$M(x) = \sum_{n \leq x} \mu(n),$$

ovvero la somma parziale x-esima della serie composta dai valori della funzione di Möbius.
Mertens congetturò che

$$|M(x)| < \sqrt{n},$$

tuttavia la sua affermazione si rivelò essere falsa (grazie al lavoro di Odlyzko e te Riele che nel 1985 trovarono controesempi ([7])).

Teorema
L'ipotesi di Riemann è equivalente a

$$M(x) = O\left(x^{\frac{1}{2}+\varepsilon}\right),$$

per ogni $\varepsilon > 0$.

Ci sono molte altre conseguenze dell'ipotesi di Riemann. Esse vanno da altre funzioni della Teoria Analitica dei Numeri (come quella dei divisori $\sigma(n)$) fino alla Teoria dei Grafi passando per le serie di Dirichlet.

CONCLUSIONE

Di certo, questo è stato un lungo viaggio e, come tale, ha preteso numerose soste intermedie.

Si sono visti vari ambiti matematici – apparentemente lontani – ma tutti attraversati da quel filo conduttore che unisce la ζ e l'ipotesi di Riemann ai numeri primi e, in generale, alla Teoria dei Numeri.
E' stata senz'altro una lunga scalata, nella quale si sono alternati sentieri quasi pianeggianti (come potrebbe essere l'introduzione alla TDN) a pareti pressoché verticali (come potrebbe essere la sezione sui teoremi di von Mangoldt), tuttavia questo viaggio è giunto al termine e siamo giunti ad abbracciare, si spera con una sufficiente completezza, l'ipotesi di Riemann e la sua importanza all'interno della matematica.
Il *Santo Graal* contenuto all'interno dell'articolo di Riemann, infatti, è proprio il collegamento tra i termini correttivi della stima di Gauss per la funzione $\pi(x)$ e gli zeri (non banali) della ζ. E' una relazione che, di per sé, lega due ambiti apparentemente lontani della matematica, quali potrebbero sembrare l'Analisi Complessa e la Teoria dei Numeri.

In realtà, però, il viaggio è solo all'inizio e forse, come per tanti altri problemi o teorie matematiche affascinanti, non avrà mai una fine o un traguardo da raggiungere. Non mi riferisco solo all'osservazione evidente che l'ipotesi di Riemann è ancora da risolvere, ma anche alle sue generalizzazioni e alla miriade di problemi che ne sono derivati.

In matematica, infatti, non esiste né un inizio né una fine. La risposta ad ogni domanda non è altro che il punto di partenza per altre domande: è così, è stato così e sarà sempre così...

... ed è proprio questo il bello della matematica._

E se proprio devo concludere citando una frase famosa (come costume per tesi di questo tipo), lascio solo un

> "Considerate la vostra semenza:
> fatti non foste per viver come bruti,
> ma per seguir virtute e canoscenza."

(D. Alighieri, *Divina Commedia, Inferno, canto XXVI*, vv. 118-120.)

APPENDICE I: ARTICOLO DI RIEMANN

In questa sezione sarà riportata in forma integrale una traduzione in italiano dell'articolo originale di Riemann. La traduzione – ad opera del sottoscritto – non è squisitamente letterale ma, laddove mi sono concesso delle libertà, esse riguardano solamente modifiche atte a fornire una maggiore scorrevolezza della lettura ([18]).
Un'autorevole traduzione dell'articolo di Riemann dall'originale tedesco all'inglese, si può anche trovare alla fine di ([9], §APPENDIX).

SUL NUMERO DI PRIMI MINORI DI UNA CERTA QUANTITA' DATA (Ueber die Anzahl del Primzahlen unter einer gegebenen Grösse)

Credo di poter comunicare meglio i miei ringraziamenti per l'onore conferitomi in qualche grado dall'Accademia ammettendomi come uno dei suoi corrispondenti, se faccio velocemente uso del permesso ricevuto per comunicare una ricerca sulla distribuzione dei numeri primi; un argomento che forse non sembra del tutto indegno per una ricerca del genere, dato l'interesse che Gauss e Dirichlet hanno mostrato molto a lungo nei suoi confronti.[1]
Per questa ricerca il mio punto di partenza è dato dall'osservazione di Eulero sul prodotto

$$\prod \frac{1}{1-\frac{1}{p^s}} = \sum \frac{1}{n^s},$$

se si sostituisce a p tutti i numeri primi e a n tutti i numeri.
La funzione di variabile complessa s rappresentata da queste due espressioni, se converge, la indico con $\zeta(s)$.[2] Entrambe le espressioni convergono solo se la parte reale di s è maggiore di 1; in contemporanea, si può trovare facilmente un'espressione per la funzione che rimane sempre valida.
Facendo uso dell'equazione

$$\int_0^\infty e^{-nx} x^{s-1} dx = \frac{\prod(s-1)}{n^s}$$

prima si vede che [3]

$$\prod(s-1)\zeta(s) = \int_0^\infty \frac{x^{s-1}dx}{e^x-1}.$$

Se ora si considera l'integrale

$$\int \frac{(-x)^{s-1}dx}{e^x-1}$$

da $+\infty$ a $+\infty$ fatto in senso positivo intorno ad un dominio che include lo zero ma senza altri punti di discontinuità dell'integrando al suo interno, allora esso si vede facilmente essere uguale a

$$\left(e^{-\pi si} - e^{\pi si}\right)\int_0^\infty \frac{x^{s-1}dx}{e^x - 1},$$

provando che, nella funzione a più valori $(-x)^{s-1} = e^{(s-1)\log(-x)}$, il logaritmo di $-x$ è determinato essere un numero reale quando x è negativo. Dunque

$$2\sin\pi s\prod(s-1)\zeta(s) = i\int_{+\infty}^{+\infty}\frac{(-x)^{s-1}dx}{e^x - 1},$$

nel quale l'integrale ha il significato appena specificato.

Questa equazione, ora, da il valore della funzione $\zeta(s)$ per tutti i numeri complessi s e mostra che questa funzione è ad un solo valore e finita per tutti i valori finiti di s ad eccezione di 1, ed anche che è zero se s è uguale ad un intero negativo pari.[4]

Se la parte reale di s è negativa, allora, invece di essere preso in senso positivo intorno al dominio specifico, questo integrale può anche essere fatto in senso negativo intorno al dominio contenente tutte le restanti quantità complesse, visto che l'integrale fatto per valori di modulo infinitamente grande è poi infinitamente piccolo. Per quanto, all'interno di questo dominio, l'integrando ha discontinuità solo se x diventa uguale a tutti i multipli di $\pm 2\pi i$, e l'integrale è così uguale alla somma degli integrali fatti in senso negativo intorno a questi valori. Ma l'integrale intorno al valore $n2\pi i$ è uguale a $(-n2\pi i)^{s-1}(-2\pi i)$, si ottiene da questa

$$2\pi\sin\pi s\prod(s-1)\zeta(s) = (2\pi)^s\sum n^{s-1}((-i)^{s-1} + i^{s-1}),$$

così una relazione tra $\zeta(s)$ e $\zeta(1-s)$, in cui, attraverso l'utilizzo delle note proprietà della funzione $\prod$, può essere espressa come segue:

$$\prod\left(\frac{s}{2} - 1\right)\pi^{-\frac{s}{2}}\zeta(s)$$

non cambia quando s è sostituito da $1 - s$.[5]

Questa proprietà della funzione mi induce ad introdurre, al posto di $\prod(s-1)$, l'integrale $\prod\left(\frac{s}{2} - 1\right)$ nel termine generale della serie $\sum\frac{1}{n^s}$, attraverso la quale uno ottiene un'espressione molto conveniente per la funzione $\zeta(s)$. Infatti [6]

$$\frac{1}{n^s}\prod\left(\frac{s}{2} - 1\right)\pi^{-\frac{s}{2}} = \int_0^\infty e^{-nn\pi x}\, x^{\frac{s}{2}-1}dx,$$

così, se si pone

$$\sum_1^\infty e^{nn\pi x} = \psi(x)$$

allora

$$\prod\left(\frac{s}{2} - 1\right)\pi^{-\frac{s}{2}}\zeta(s) = \int_0^\infty \psi(x)x^{\frac{s}{2}-1}dx,$$

o da

$$2\psi(x) + 1 = x^{-\frac{1}{2}}\left(2\psi\left(\frac{1}{x}\right) + 1\right),$$

$$\prod\left(\frac{s}{2} - 1\right)\pi^{-\frac{s}{2}}\zeta(s) = \int_1^\infty \psi(x)x^{\frac{s}{2}-1}dx + \int_1^\infty \psi\left(\frac{1}{x}\right)x^{\frac{s-3}{2}}dx + \frac{1}{2}\int_0^1\left(x^{\frac{s-3}{2}} - x^{\frac{s}{2}-1}\right)dx$$

$$= \frac{1}{s(s-1)} + \int_1^\infty \psi(x)\left(x^{\frac{s}{2}-1} + x^{-\frac{1+s}{2}}\right)dx.$$

Io ora pongo $s = \frac{1}{2} + ti$ e [7]

$$\prod\left(\frac{s}{2}\right)(s-1)\pi^{-\frac{s}{2}}\zeta(s) = \xi(t)$$

e allora

$$\xi(t) = \frac{1}{2} - \left(tt + \frac{1}{4}\right)\int_1^\infty \psi(x)x^{-\frac{3}{4}}\cos\left(\frac{1}{2}t\log x\right)dx$$

o, in aggiunta,

$$\xi(t) = 4\int_1^\infty \frac{d\left(x^{\frac{3}{2}}\psi'(x)\right)}{dx}x^{-\frac{1}{4}}\cos\left(\frac{1}{2}t\log x\right)dx.$$

Questa funzione è finita per tutti i valori finiti di t, e permette essa stessa di essere sviluppata in potenze di tt con una serie convergente molto rapidamente.[8] Dal fatto che, per un valore di s la cui parte reale è più grande di 1, $\log\zeta(s) = -\sum\log(1-p^{-s})$ resta finito, e dal fatto che la stessa cosa vale per i logaritmi degli altri fattori di $\xi(t)$, segue che la funzione $\xi(t)$ può solo annullarsi se la parte immaginaria di t si trova tra $\frac{1}{2}i$ e $-\frac{1}{2}i$.[9] Il numero di radici di $\xi(t) = 0$, la cui parte reale è tra 0 e T è approssimativamente uguale a [10]

$$\frac{T}{2\pi}\log\frac{T}{2\pi} - \frac{T}{2\pi};$$

poiché l'integrale $\int d\log\xi(t)$, fatto in senso positivo intorno alla regione consistente tra i valori di t la cui parte immaginaria si trova $\frac{1}{2}i$ e $-\frac{1}{2}i$ e la cui parte reale sta tra 0 e T, è (fino ad una frazione dell'ordine di grandezza della quantità $\frac{1}{T}$) uguale a $\left(T\log\frac{T}{2\pi} - T\right)i$; questo integrale comunque è uguale al numero delle radici di $\xi(t) = 0$ che giacciono in questa regione, moltiplicati per $2\pi i$. Ora, si trova in effetti approssimativamente questo numero di radici reali entro questi limiti, ed è molto probabile che tutte le radici sono reali.[11] Certamente ci si augura una piccola dimostrazione qui; nel frattempo io ho temporaneamente messo da parte la ricerca per questo dopo qualche futile tentativo di sfuggita, così come sembra non necessario per il prossimo obiettivo della mia indagine.

Se si indicano con α tutte le radici dell'equazione $\xi(\alpha) = 0$, si può esprimere $\log\xi(t)$ come[12]

$$\sum\log\left(1 - \frac{tt}{\alpha\alpha}\right) + \log\xi(0);$$

e, dal fatto che la densità delle radici della quantità t cresce con t solo come $\log\frac{t}{2\pi}$, segue che questa espressione converge e diventa per un t infinito solo infinita come $t\log t$; così differisce dal $\log\xi(t)$ per una funzione di tt, che per un t finito resta continua e finita e, quando è divisa da tt, diventa infinitamente piccola per t infinito. Questa differenza è, di conseguenza, una costante, il cui valore può essere determinato ponendo $t = 0$.

Grazie a questi metodi, il numero dei primi che sono più piccoli di x può essere determinato.

Sia $F(x)$ uguale a questo numero quando x non è esattamente uguale ad un numero primo; ma sia più grande di $\frac{1}{2}$ quando x sia un numero primo, allora, per ogni x a cui qui c'è un salto nel valore di $F(x)$,[13]

$$F(x) = \frac{F(x+0) + F(x-0)}{2}.$$

Se nell'identità

$$\log\zeta(s) = -\sum\log(1-p^s) = \sum p^{-s} + \frac{1}{2}\sum p^{-2s} + \frac{1}{3}\sum p^{-3s} + \cdots$$

adesso si sostituisce

$$p^{-s} \text{ con } s\int_p^\infty x^{-s-1}dx, \quad p^{-2s} \text{ con } s\int_{p^2}^\infty x^{-s-1}dx, \qquad \ldots,$$

uno ottiene

$$\frac{\log\zeta(s)}{s} = \int_1^\infty f(x)x^{-s-1}dx,$$

se uno indica

$$F(x) + \frac{1}{2}F\left(x^{\frac{1}{2}}\right) + \frac{1}{3}F\left(x^{\frac{1}{3}}\right) + \cdots$$

tramite $f(x)$.[14]

Questa equazione è valida per ogni numero complesso $a + bi$ di s nel quale $a > 1$. Se, però, l'equazione

$$g(s) = \int_0^\infty h(x)x^{-s}d\log x$$

è valida in mezzo a questo intervallo, allora, facendo uso del teorema di Fourier, si può esprimere la funzione h in termini della funzione g. L'equazione si decompone, se $h(x)$ è reale e

$$g(a+bi) = g_1(b) + ig_2(b),$$

nelle seguenti due:

$$g_1(b) = \int_0^\infty h(x)x^{-a}\cos(b\log x)\,d\log x,$$

$$ig_2(b) = -i\int_0^\infty h(x)x^{-a}\sin(b\log x)\,d\log x.$$

Se uno moltiplica entrambe le equazioni con

$$(\cos(b\log y) + i\sin(b\log y))db$$

e li integra da $-\infty$ a $+\infty$, allora uno ottiene $\pi h(y)y^{-\alpha}$ a destra di entrambe, in accordo dai teoremi di Fourier; così, se si aggiunge entrambe le equazioni e le moltiplica per iy^α, uno ottiene

$$2\pi i h(y) = \int_{a-\infty i}^{a+\infty i} g(s)y^s ds,$$

dove l'integrazione è svolta affinché la parte reale di s resti costante.

Per un valore di y in cui c'è un salto nel valore di $h(y)$, l'integrale porta alla media dei valori della funzione h ad ogni lato del salto. Dal modo in cui la funzione f è definita, vediamo che essa ha la stessa proprietà, e, in generale, [15]

$$f(y) = \frac{1}{2\pi i}\int_{a-\infty i}^{a+\infty i}\frac{\log\zeta(s)}{s}y^s\,ds.$$

Si può sostituire per $\log\zeta$ l'espressione

$$\frac{s}{2}\log\pi - \log(s-1) - \log\prod\left(\frac{s}{2}\right) + \sum^\alpha\log\left(1 + \frac{\left(s-\frac{1}{2}\right)^2}{\alpha\alpha}\right) + \log\xi(0)$$

trovata prima; comunque gli integrali dei termini individuali di questa espressione non convergono, quando sono estesi all'infinito, per la ragione che è appropriato convertire la precedente equazione tramite l'integrazione per parti in

$$f(x) = -\frac{1}{2\pi i}\frac{1}{\log x}\int_{a-\infty i}^{a+\infty i} \frac{d\,\frac{\log\zeta(s)}{s}}{ds} x^s ds$$

Da

$$-\log\prod\left(\frac{s}{2}\right) = \lim\left(\sum_{n=1}^{n=m} \log\left(1+\frac{s}{2n}\right) - \frac{s}{2}\log m\right),$$

per $m = \infty$ e pertanto

$$-\frac{d\,\frac{1}{s}\log\prod\left(\frac{s}{2}\right)}{ds} = \sum_{1}^{\infty} \frac{d\,\frac{1}{s}\log\left(1+\frac{s}{2n}\right)}{ds},$$

segue poi che tutti i termini dell'espressione di $f(x)$, ad eccezione di

$$\frac{1}{2\pi i}\frac{1}{\log x}\int_{a-\infty i}^{a+\infty i} \frac{1}{ss}\log\xi(0)\, x^s ds = \log\xi(0),$$

hanno la forma

$$\pm\frac{1}{2\pi i}\frac{1}{\log x}\int_{a-\infty i}^{a+\infty i} \frac{d\left(\frac{1}{s}\log\left(1-\frac{s}{\beta}\right)\right)}{ds} x^s ds.$$

Ma ora

$$\frac{d\left(\frac{1}{s}\log\left(1-\frac{s}{\beta}\right)\right)}{d\beta} = \frac{1}{(\beta-s)\beta},$$

e, se la parte reale di s è più grande della parte reale di β,

$$-\frac{1}{2\pi i}\int_{a-\infty i}^{a+\infty i} \frac{x^s ds}{(\beta-s)\beta} = \frac{x^\beta}{\beta} = \int_{\infty}^{x} x^{\beta-1}dx,$$

o

$$= \int_{0}^{x} x^{\beta-1}dx,$$

dipende se la parte reale di β è negativa o positiva. Si ha come risultato

$$\frac{1}{2\pi i}\frac{1}{\log x}\int_{a-\infty i}^{a+\infty i} \frac{d\left(\frac{1}{s}\log\left(1-\frac{s}{\beta}\right)\right)}{ds} x^s ds = -\frac{1}{2\pi i}\int_{a-\infty i}^{a+\infty i} \frac{1}{s}\log\left(1-\frac{s}{\beta}\right) x^s ds$$
$$= \int_{\infty}^{x} \frac{x^{\beta-1}}{\log x}dx + \text{cost}$$

nella prima e

$$= \int_{0}^{x} \frac{x^{\beta-1}}{\log x}dx + \text{cost.}$$

nel secondo caso.

Nel primo caso la costante di integrazione è determinata se uno lascia che la parte reale di β diventa infinitamente negativa; nel secondo caso l'integrale tra 0 e x porta a valori separati da $2\pi i$, che dipendono dal fatto che l'integrazione è fatta lungo valori complessi con argomento positivo o negativo, e diventa infinitamente piccolo, per precedenti cammini, dove il

coefficiente di i nel valore di β diventa infinitamente positivo, o per quest'ultimo, quando questo coefficiente diventa infinitamente negativo. Da questo si è visto come a sinistra $\log\left(1-\frac{s}{\beta}\right)$ è da determinare in modo che la costante di integrazione sparisce.

Attraverso l'inserimento di questi valori nell'espressione di $f(x)$ uno ottiene [16]

$$f(x) = Li(x) - \sum^{\alpha}\left(Li\left(x^{\frac{1}{2}+\alpha i}\right) + Li\left(x^{\frac{1}{2}-\alpha i}\right)\right) + \int_x^{\infty} \frac{1}{x^2-1}\frac{dx}{x\log x} + \log\xi(0),$$

se nel $\sum^{\alpha}$ uno sostituisce per α tutte le radici positive (o le radici che hanno una parte reale positiva) dell'equazione $\xi(\alpha) = 0$, ordinate per grandezza. Si può facilmente mostrare, per mezzo di più discussioni attraverso la funzione ξ, che con questo ordine di termini il valore della serie

$$\sum\left(Li\left(x^{\frac{1}{2}+\alpha i}\right) + Li\left(x^{\frac{1}{2}-\alpha i}\right)\right)\log x$$

è in accordo con il valore limite al quale

$$\frac{1}{2\pi i}\int_{a-bi}^{a+bi} \frac{d\frac{1}{s}\sum\log\left(1+\frac{\left(s-\frac{1}{2}\right)^2}{\alpha\alpha}\right)}{ds} x^s ds$$

converge quando la quantità b cresce senza limite; comunque dove riordinata essa può assumere ogni qualsiasi valore reale arbitrario.

Da $f(x)$ uno ottiene $F(x)$ tramite l'inversione della relazione

$$f(x) = \sum \frac{1}{n} F\left(x^{\frac{1}{n}}\right),$$

per ricavare l'equazione

$$F(x) = \sum(-1)^{\mu}\frac{1}{m} f\left(x^{\frac{1}{m}}\right),$$

nella quale si sostituisce alla m la serie che consiste in quei numeri naturali che non sono divisibili da nessun quadrato a parte 1, e nel quale μ indica il numero di fattori primi di m.[17]

Se uno restringe $\sum^{\alpha}$ ad un finito numero di termini, allora la derivata dell'espressione per $f(x)$ o, fino a una parte diminuisce molto rapidamente al crescere di x,

$$\frac{1}{\log x} - 2\sum^{\alpha}\frac{\cos(\alpha\log(x))\, x^{-\frac{1}{2}}}{\log x}$$

dà un'espressione approssimata della densità dei numeri primi + metà della densità dei quadrati dei numeri primi + un terzo della densità dei cubi dei numeri primi etc. alla grandezza x.[18]

La nota espressione approssimante $F(x) = Li(x)$ è pertanto valida fino a quantità dell'ordine $x^{\frac{1}{2}}$ e dà un valore piuttosto grande; perché i termini non periodici nell'espressione per $F(x)$ sono, a parte quantità che non crescono all'infinito con x:[19]

$$Li(x) - \frac{1}{2}Li\left(x^{\frac{1}{2}}\right) - \frac{1}{3}Li\left(x^{\frac{1}{3}}\right) - \frac{1}{6}Li\left(x^{\frac{1}{6}}\right) - \frac{1}{7}Li\left(x^{\frac{1}{7}}\right) + \cdots$$

In effetti, nella comparazione tra $Li(x)$ con il numero dei numeri primi minori di x, intrapresa da Gauss e Goldschmidt e portata avanti fino a $x =$ tre milioni, questo numero si è mostrato essere, nelle prime centinaia di migliaia, sempre minore di $Li(x)$; infatti la differenza cresce,

con molte fluttuazioni, gradualmente con x. Ma anche il crescere e il decrescere nella densità dei primi da un posto all'altro che dipende dai termini periodici ha già suscitato attenzioni, senza comunque essere stata osservata nessuna legge che governi questo comportamento. In ogni futuro conto sarà interessante tenersi al passo con l'influenza degli individuali termini periodici nell'espressione della densità dei numeri primi. Un comportamento più regolare rispetto a quello di $F(x)$ potrebbe essere mostrato dalla funzione $f(x)$, che già nelle prime centinaia si è già mostrata molto distintamente essere mediamente in accordo con $Li(x) + \log \xi(0)$.[20]

NOTE

1. Riemann ringrazia l'accademia per l'onore che le ha dato nell'ammetterlo come suo corrispondente e decide di ripagarli (in un certo qual modo) dedicando loro la sua ricerca sui numeri primi. L'articolo è in prima persona singolare salvo espressioni come "uno ha" o "uno ottiene" per quanto riguarda i calcoli. Nella mia tradizione ho scelto di convertire queste espressioni negli usuali "si ha" o "si ottiene".
2. Si tratta della definizione della ζ come estensione della serie armonica generalizzata e della rappresentazione della stessa trovata grazie alla formula del prodotto di Eulero (§10.1.6). Riemann, da parte sua, specifica che "convergono per $Re(s) > 1$" e – laddove convergono – "indicherà queste espressioni con $\zeta(s)$".
3. Come detto nella sezione dedicata alla funzione Γ (§8.4), Riemann si serve della funzione $\prod$, che, in alcuni casi, consente anche di semplificare i calcoli (§14.3).
   In queste righe troviamo il prolungamento della ζ all'intero piano complesso: in particolar modo il matematico tedesco inizia ricordando la rappresentazione integrale della ζ stessa (§11.3).
4. Riemann sta per dare la *prima prova* (delle due che si possono trovare nel suo articolo) dell'equazione funzionale della ζ (§12.3.1), tramite la rappresentazione integrale della stessa citata nella nota precedente.
   Fa notare, inoltre, che $\zeta(s) = 0$, per s intero negativo pari: come abbiamo visto (§12.3.2), questi sono i così detti zeri banali causati dall'annullamento del seno – presente nell'equazione funzionale – dovuto al suo argomento.
5. Riemann finisce di dare la sua *prima prova* dell'equazione funzionale della ζ (§12.3.1).
6. Quella che segue è la *seconda prova* dell'equazione funzionale della ζ, ottenuta servendosi delle funzioni di Jacobi e delle loro proprietà (§12.3.4).
   Possiamo notare che Riemann si serve di una scrittura inusuale. A parte la funzione $\prod$ in luogo della Γ, compaiono spesso nn in luogo di n^2, $\alpha\alpha$ in luogo di α^2, ecc...
7. Riemann pone
   $$s = \frac{1}{2} + ti,$$
   nel suo caso, però, t è complesso, non reale: la scrittura usuale, infatti, presuppone $t = Re(s)$. Una sostituzione del genere, però, consente a Riemann di formulare la sua ipotesi con un "elegante"

   > "è probabile che tutte le radici di ξ siano reali",

   quando invece l'ipotesi di Riemann universalmente accettata parla di $Re(s) = \frac{1}{2}$ (§13.3). Riferendoci a questa insolita sostituzione scriveremo "la *sua* ξ" altrimenti intenderemo il generico termine " ξ " in luogo della formulazione della ξ universalmente accettata.
8. Come detto (§13.1.5), Riemann non fa nessun cenno a questa "convergenza molto rapida".

9. Riemann "dimostra" che gli zeri della *sua* ξ (come detto ha posto $s = \frac{1}{2} + it$, con t complesso) sono tali che $-\frac{1}{2} < Re(s) < \frac{1}{2}$. In termini moderni equivale a dire che gli zeri della ξ, cioè quelli non banali della ζ, sono tali che $0 < Re(s) < 1$.
10. Dando anche qualche cenno di dimostrazione, Riemann enuncia la densità degli zeri lungo la striscia critica (§14.1.4).
11. "… è molto probabile che tutte le radici siano reali": questa è l'ipotesi di Riemann (§13.3.2).
    Il matematico tedesco *ipotizza* che gli zeri della *sua* ξ siano reali, cioè che gli zeri della ξ siano tali che $Re(s) = 1/2$. Prosegue dicendo che "ci si augura una piccola dimostrazione qui" ma lui ha "temporaneamente messo da parte la ricerca per questo dopo qualche futile tentativo di sfuggita, così come sembra non necessario per il prossimo obiettivo della mia indagine".
    Si potrebbe pensare a un'analogia con l'ultimo teorema di Fermat e con il fatto che Fermat scrisse di "avere una meravigliosa dimostrazione di questo fatto [cioè che $x^n + y^n = z^n$ non ha radici intere per $n > 2$, *n.d.A.*] ma che avesse poco spazio a disposizione per riportarla". Quella dell'ultimo teorema di Fermat è stata a lungo pensata come una beffa (§6.1.7), tuttavia nel 1995 Wiles riuscì a dimostrarlo (con metodi tutt'altro che elementari). In questo caso, però, le parole di Riemann sembrerebbero più veritiere poiché il suo scopo, nell'articolo di ricerca, non è quello di dimostrare che la *sua* ξ ha zeri reali, ma quello di trovare una formula *esatta* per il calcolo di $\pi(x)$ come dice il titolo stesso.
12. La formulazione di Riemann è corretta anche se raramente utilizzata. Ci si serve dell'usuale
    $$\log(\xi(s)) = \log(\xi(0)) + \sum_{\rho} \log\left(1 - \frac{s}{\rho}\right).$$
    La rappresentazione per mezzo di
    $$\sum_{\rho} \log\left(1 - \frac{tt}{\alpha\alpha}\right),$$
    deriva dal riformulare la ξ isolando l'asse di simmetria $Re(s) = 1/2$ considerando, nella *sua* ξ, il solito $s = \frac{1}{2} + it$ con t complesso (per mezzo di calcoli più complicati di quelli usuali). Ricordiamo, sempre, che "tt" vale "t^2" e "$\alpha\alpha$" vale "α^2" e che, in questo calcolo, la *sua* $\xi(0)$ equivale a $\xi\left(\frac{1}{2}\right)$.
13. Riemann introduce la sua $\pi(x)$, indicandola con $F(x)$. La differenza con l'usuale $\pi(x)$ sta solo nel passaggio ad un n primo, nel quale $F(x)$ assume un valore intermedio tra il precedente e il successivo (§16.2).
14. Riemann passa ad introdurre la funzione $J(x)$, indicandola con un generico $f(x)$.
15. Dopo aver introdotto la $J(x)$, Riemann si serve dell'inversione di Fourier (chiamandola "teorema di Fourier"), per esplicitarla in termini della ζ.
16. Si è appena conclusa la parte più "spiegata" dell'articolo di Riemann, quella che testimonia gli sforzi profusi dal matematico tedesco nel trovare una formula per la $J(x)$ (§16.5). Riemann, inoltre, utilizza le scritture

$$\frac{1}{2} + \alpha i, \qquad \frac{1}{2} - \alpha i,$$

in luogo dei "moderni"

$$\rho, \qquad 1 - \rho,$$

per indicare gli zeri della ξ. La sostanza, però, è la stessa.

17. Tramite l'inversione di Möbius, Riemann trova la formula esatta per la $\pi(x)$ a partire dalla definizione della $J(x)$ (§16.6).
18. Riemann accenna al fatto che la serie, in realtà, è finita e che la si può troncare ottenendo buone approssimazioni.
19. Riemann spiega che questo risultato non è in contrasto con l'espressione trovata da Gauss (dell'approssimazione di $\pi(x)$ con $Li(x)$ (§16.7)) ma che, anzi, essa ne è una conferma.
20. Riemann conclude il suo articolo con le evidenze numeriche (relative al suo tempo) per l'approssimazione di Gauss

    $$\pi(x) \sim Li(x),$$

    e per la sua formula dei primi. Accenna anche al fatto che tale approssimazione oscillerà per difetto e per eccesso: cosa che si può riscontrare nella scrittura della sua formula che si serve della funzione μ (§16.7).

    Conclude esprimendo la speranza che calcoli futuri gli diano ragione: fino ad ora gli danno ragione in quanto non si è ancora trovato uno zero "fuori posto" oppure troppa differenza tra la sua approssimazione della $\pi(x)$ e i valori veri della stessa.

APPENDICE II: IL TEOREMA DI HARDY

In questa sezione discuteremo brevemente un notevole risultato di Hardy circa l'infinità degli zeri della funzione ξ (quindi quelli non banali della ζ) lungo la linea critica $Re(s) = 1/2$.
Tale risultato è il capostipite di altri teoremi successivi di questo tipo che trattano dell'andamento asintotico e della percentuale degli zeri lungo la linea critica rispetto a quelli totali.

Teorema (Hardy)
Esistono infiniti zeri ρ per la funzione ξ tali che $Re(\rho) = 1/2$.
Dimostrazione
Discuteremo, nei caratteri essenziali, la dimostrazione offerta da Titchmarsh in ([25]), che fa uso della formula di Riemann-Siegel. Ci sono, infatti, diverse dimostrazioni di questo teorema che, generalmente, fanno uso delle trasformate di Fourier e di altri risultati avanzati riguardo l'integrazione complessa (per es. ([26], §10)).
Richiamiamo la formula di Riemann-Siegel (§15.2)

$$\begin{aligned}\zeta\left(\frac{1}{2}+it\right) &= \Xi(t)\\ &= 2\sum_{n=1}^{N}\frac{cos(\theta(t)-t\log(n))}{n^{1/2}}\\ &+\frac{e^{i\theta(t)}e^{-t\pi/2}}{(2\pi)^{1/2+it}e^{-t\pi/4}(1-ie^{t\pi})}\int_{C_N}\frac{(-x)^{-1/2+it}e^{-Nx}dx}{e^{x}-1},\end{aligned}$$

essa si può riscrivere, in maniera approssimata, come

$$\zeta\left(\frac{1}{2}+it\right) = \Xi(t) = 2\sum_{n\leq\sqrt{t/(2\pi)}}\frac{cos(\theta(t)-t\log(n))}{n^{1/2}} + O\left(t^{-1/4}\right),$$

ricordando che N è un intero positivo tale che C_N contiene i punti $\pm 2n\pi i$, per $n = 0, \dots, N$ (si confronti con quanto detto per il cammino di integrazione utilizzato nella prima dimostrazione dell'equazione funzionale della ζ in (§12.3.1, Figura 12.6)) mentre l'ultima parte è l'ordine dell'errore che si commette considerando solamente la sommatoria.
In questa scrittura $\theta(t)$ era la funzione di Riemann-Siegel definita come

$$\theta(t) = \arg\left(\Gamma\left(\frac{2it+1}{4}\right)\right) - \frac{\log(\pi)}{2}t,$$

la quale ha anche la seguente espansione asintotica ([26])

$$\theta(t) \sim \frac{t}{2}\log\left(\frac{t}{2\pi}\right) - \frac{t}{2} - \frac{\pi}{8} + \frac{1}{48t} + \frac{t}{5760t^3} + O(t^{-5}).$$

Si può, inoltre, provare che le prime due derivate – per t sufficientemente grande – sono circa

$$\theta'(t) \sim \frac{t}{2}\log\left(\frac{t}{2\pi}\right), \qquad \theta''(t) \sim \frac{1}{2t},$$

da cui si deduce che $\theta(t)$ è monotona crescente e che un'eventuale equazione

$$\theta(t) = \nu\pi,$$

ha l'unica soluzione

$$t_\nu \sim \frac{2\pi\nu}{\log(\nu)}.$$

La funzione $\Xi(t)$, valutata per $t = t_\nu$ diventa

$$\Xi(t_\nu) = (-1)^\nu 2g(t_\nu) + O\left(t_\nu^{-1/4}\right),$$

nel quale si è posto

$$g(t_\nu) = \sum_{n \le \sqrt{t_\nu/(2\pi)}} \frac{\cos(t_\nu\, log(n))}{n^{1/2}} = 1 + \frac{\cos(t_\nu\, log(2))}{2^{1/2}} + \cdots,$$

che ci dice che la quantità $g(t_\nu)$ è uguale a 1 più una serie di termini che cambiano di segno e decrescono in valore assoluto e, conseguentemente, si può appurare che $g(t_\nu) > 0$ e che $\Xi(t_\nu)$ cambia segno nell'intervallo $]t_\nu, t_\nu + 1[$.

Il punto della dimostrazione è far vedere che

$$\sum_{\nu=M+1}^{N} \Xi(t_{2\nu}) \sim 2N, \qquad \sum_{\nu=M+1}^{N} \Xi(t_{2\nu+1}) \sim -2N.$$

In questo modo, facendo tendere $N \to \infty$, possiamo concludere che $\Xi(t)$ cambia segno infinite volte – cioè che $\Xi(t)$ ha infiniti zeri di molteplicità dispari – perché $\Xi(t_{2\nu})$ è positiva infinite volte e $\Xi(t_{2\nu+1})$ è negativa infinite volte.

Consideriamo, ora, la somma

$$\sum_{\nu=M+1}^{N} g(t_{2\nu}) = \sum_{\nu=M+1}^{N} \left(\sum_{n \le \sqrt{t_{2\nu}/(2\pi)}} \frac{\cos(t_{2\nu} \log(n))}{n^{1/2}} \right)$$

$$= N - M + \sum_{2 \le n \le \sqrt{t_{2\nu}/(2\pi)}} \frac{1}{n^{1/2}} \sum_{\tau \le t_{2\nu} \le t_{2N}} \cos(t_{2\nu} \log(n)),$$

dove $\tau = \max\{2\pi n^2, \tau_{2M+2}\}$. Indicando

$$\phi(\nu) = \frac{t_{2\nu} \log(n)}{2\pi},$$

l'ultima sommatoria precedente diventa semplicemente

$$\sum_{\tau \le t_{2\nu} \le t_{2N}} \cos(t_{2\nu} \log(n)) = \sum_{\tau \le t_{2\nu} \le t_{2N}} \cos\left(2\pi\phi(\nu)\right).$$

Possiamo notare che dal fatto che $\theta(t_{2\nu}) = 2\nu\pi$, otteniamo

$$\theta'(t_{2\nu}) \frac{dt_{2\nu}}{d\nu} = 2\pi,$$

e troviamo

$$\phi'(\nu) = \frac{\log(n)}{2\pi} \frac{dt_{2\nu}}{d\nu} = \frac{\log(n)}{\theta'(t_{2\nu})} > 0.$$

Allo stesso tempo, per ν grande

$$\phi''(\nu) = -\log(n) \frac{\theta''(t_{2\nu})}{[\theta'(t_{2\nu})]^2} \frac{dt_{2\nu}}{d\nu} < -\frac{8\pi \log(n)}{t_{2\nu} \log^3(t_{2\nu})} < -A \frac{\log(n)}{t_{2N} \log^3(t_{2N})},$$

per A costante opportuna.

Si può provare che se $f \in C^2$ nell'intervallo $[a, b]$, con $\lambda \le -f''(x) \le h\lambda$ in $[a, b]$ con $b \ge a + 1$, allora e $\lambda > 0$ costante reale,

$$\sum_{a<n\leq b} e^{2\pi i f(n)} = O\left[(b-a)\sqrt{\lambda}\right] + O\left(\frac{1}{\sqrt{\lambda}}\right),$$

e, come conseguenza, si ha

$$\sum_{\tau\leq t_{2\nu}\leq t_{2N}} \cos(t_{2\nu}\log(n)) = \frac{e^{2\pi i\phi(\nu)} + e^{-2\pi i\phi(\nu)}}{2}$$

$$= O\left(t_{2N}\frac{\log^{1/2}(n)}{\sqrt{t_{2N}}\log^{3/2}(t_{2N})}\right) + O\left(\frac{\sqrt{t_{2N}}\log^{3/2}(t_{2N})}{\log^{1/2}(n)}\right) = O\left(\sqrt{t_{2N}}\log^{\frac{3}{2}}(t_{2N})\right).$$

Allora

$$\sum_{2\leq n\leq\sqrt{t_{2\nu}/(2\pi)}} \frac{1}{n^{1/2}} \sum_{\tau\leq t_{2\nu}\leq t_{2N}} \cos(t_{2\nu}\log(n)) = O\left(t_{2N}^{3/4}\log^{3/2}(t_{2N})\right) = O\left(N^{3/4}\log^{3/2}(N)\right),$$

così abbiamo provato la seguente equazione:

$$\sum_{\nu=M+1}^{N} \Xi(t_{2\nu}) = 2N + O\left(N^{3/4}\log^{3/2}(N)\right) \sim 2N.$$

In un modo simile, si può provare anche che

$$\sum_{\nu=M+1}^{N} \Xi(t_{2\nu+1}) \sim -2N,$$

in modo da dimostrare il teorema.

Osservazioni

Come detto, il teorema venne dimostrato da Hardy servendosi di tecniche avanzate di analisi complessa; in questa sezione abbiamo riportato cenni sul procedimento utilizzato, in seguito, da Titchmarsh per giungere alla medesima conclusione.

Questo risultato è il primo di una serie simile di risultati atti ad indagare sulla quantità di zeri che giacciono lungo la linea critica.

Hardy e Littlevood, nel 1921, provarono che "per T sufficientemente grande, nel segmento $\left[\frac{1}{2}, \frac{1}{2} + iT\right]$ ci sono almeno KT zeri, con K costante positiva".

In seguito, Selberg dimostrò che "il numero di tali radici della ξ lungo la linea critica è almeno $KT\log(T)$, per T sufficientemente grande e K costante opportuna".

Queste dimostrazioni si possono trovare in ([26]), ma anche in ([9], §11).

La questione, però, è tanto semplice quanto negativa: il fatto che ci sono infiniti zeri lungo la linea critica non implica che ce ne possano essere altri in zone della restante striscia critica e, oltretutto, che ce ne possano essere infiniti.

Dunque l'ipotesi di Riemann è ben lontana dall'essere risolta.

APPENDICE III: IL TEOREMA DEI NUMERI PRIMI

In questa sezione daremo una dimostrazione (abbastanza elementare) del teorema dei numeri primi e parleremo delle implicazioni che ha sui numeri primi stessi e sulle funzioni di Chebyshev.
Dimostreremo i risultati principali – a cominciare ovviamente da quelli riguardanti il Teorema dei Numeri Primi – mentre per gli altri l'esposizione sarà limitata al solo enunciato poiché secondari, complicati o non influenti per la finalità di questa sezione.
Prima di inoltrarci negli obiettivi appena dichiarati, richiamiamo brevemente le due funzioni di Chebyshev viste nella sezione di Teoria Analitica dei Numeri (§10.2.4). Avevamo, infatti, definito, per $x > 0$ reale:

- la funzione ψ di Chebyshev mediante la seguente formula
$$\psi(x) = \sum_{n \leq x} \Lambda(n)\,;$$
- la funzione ϑ di Chebyshev nel modo che segue
$$\vartheta(x) = \sum_{p \leq x} \log(p)\,, \qquad p \text{ primo.}$$

Lemma
Per la funzione di Chebyshev $\psi(x)$ si ha
$$\psi(x) = \sum_{p \leq x} \left\lfloor \frac{\log(x)}{\log(p)} \right\rfloor \log(p)\,, \qquad p \text{ primo}, \qquad x \geq 2.$$
Dimostrazione
Come abbiamo appena ricordato
$$\psi(x) = \sum_{n \leq x} \Lambda(n),$$
nella quale la Λ denota la funzione di Van Mangoldt.
Sia, dunque, $p \leq x$ primo e p^m (m intero positivo) la massima potenza di p che non supera x, cioè
$$p^m \leq x < p^{m+1}.$$
Allora
$$m \log(p) \leq \log(x) < (m+1)\log(p)$$
dunque
$$m = \left\lfloor \frac{\log(x)}{\log(p)} \right\rfloor.$$
A questo punto, per $n = p^1, p^2, \ldots, p^m$ si ha ogni volta $\Lambda(n) = \log(p)$ e perciò, in $\psi(x)$, il contributo di queste potenze vale

$$\left\lfloor \frac{\log(x)}{\log(p)} \right\rfloor \log(p).$$

La formula appena mostrata vale per $p \leq x$ primo fissato; estendendola a tutti i primi $\leq x$ che compaiono nella sommatoria abbiamo la tesi

$$\psi(x) = \sum_{p \leq x} \left\lfloor \frac{\log(x)}{\log(p)} \right\rfloor \log(p)\,, \qquad p \text{ primo}.$$

Teorema ([10], §22.2)
Per $x > 0$, vale la seguente relazione tra le funzioni ψ e ϑ di Chebyshev

$$\psi(x) = \vartheta(x) + O\left(x^{\frac{1}{2}} \log^2(x)\right).$$

Dimostrazione
La dimostrazione segue immediatamente dal teorema (§10.2.4)

$$\frac{\psi(x)}{x} - \frac{\vartheta(x)}{x} \leq \frac{\log^2(x)}{2\sqrt{x}\log(2)},$$

cioè

$$\frac{\psi(x)}{x} \leq \frac{\vartheta(x)}{x} + \frac{\log^2(x)}{2\sqrt{x}\log(2)}.$$

Moltiplicando per x ($x \geq 2$ per ipotesi), otteniamo la tesi ricordando che nella definizione di O, le costanti sono ininfluenti (§1.3.4).

Teorema ([10], §22.2)
Per ogni $n \geq 1$ intero, $\vartheta(n) < 2n\log(2)$.
Dimostrazione
Consideriamo, preliminarmente, il seguente coefficiente (intero)

$$M = \binom{2m+1}{m} = \frac{(2m+1)!}{m!\,(m+1)!} = \frac{(2m+1)\cdot(2m)\cdot\ldots\cdot(m+2)}{m!}.$$

Sviluppando binomialmente l'espansione di $(1+1)^{2m+1}$, si noterà che M comparirà due volte in tale espansione, dunque

$$2M < (1+1)^{2m+1} = 2^{2m+1},$$

cioè

$$M < 2^{2m}.$$

Ora, per ogni p primo, con $m+1 < p \leq 2m+1$, p divide il numeratore di M ma non il suo denominatore ($p \geq m$), dunque

$$\left(\prod_{m+1<p\leq 2m+1} p\right) \Bigg| M,$$

allora

$$\vartheta(2m+1) - \vartheta(m+1) = \sum_{m+1<p\leq 2m+1} \log(p) = \log\left(\prod_{m+1<p\leq 2m+1} p\right) \leq \log(M)$$
$$< 2m\log(2).$$

Ora, il teorema è banalmente vero per $n = 1,2$: supponiamolo vero per tutti gli $n \leq n_0 - 1$. Se n_0 è pari, abbiamo

$$\vartheta(n_0) = \vartheta(n_0 - 1) < 2(n_0 - 1)\log(2) < 2n_0 \log(2).$$

Se n_0 è dispari, $n_0 = 2m + 1$ per m opportuno, dunque

$$\begin{aligned}\vartheta(n_0) = \vartheta(2m+1) &= \vartheta(2m+1) - \vartheta(m+1) + \vartheta(m+1) \\ &< 2m\log(2) + 2(m+1)\log(2) = 2(2m+1)\log(2) = 2n_0\log(2),\end{aligned}$$

la sostituzione $\vartheta(m+1) < 2(m+1)\log(2)$ è dovuta al fatto che $m + 1 < n_0$ e che abbiamo supposto la proprietà vera per $n \leq n_0 - 1$. La tesi del teorema segue per induzione.

Questo teorema può essere esteso al caso di x reale positivo. Infatti, tenendo conto che $\lfloor x \rfloor$ è un intero positivo per $x > 1$,

$$\vartheta(x) = \vartheta(\lfloor x \rfloor) < 2\lfloor x \rfloor \log(2) \leq 2x\log(2),$$

cioè $\vartheta(x) < Ax$, per A un'opportuna costante reale (nel nostro caso $2\log(2)$).

Teorema ([10], §22.2)
Per n naturale,

$$n! = \prod_{p \text{ primo}} p^{j(n,p)},$$

dove per ogni primo p

$$j(n,p) = \sum_{m \geq 1} \left\lfloor \frac{n}{p^m} \right\rfloor.$$

Osservazione. La somma al secondo membro è finita poiché esiste un indice m_0 tale per cui $p^{m_0} \leq n < p^{m_0+1}$ e, per la parte intera, questo si traduce con $\left\lfloor \frac{n}{p^m} \right\rfloor = 0$ per ogni $m \geq m_0 + 1$.

Dimostrazione
Basta notare che tra i numeri $2, \ldots, n$ ci sono $\left\lfloor \frac{n}{p} \right\rfloor$ multipli di p, $\left\lfloor \frac{n}{p^2} \right\rfloor$ multipli di p^2 e così via. Inoltre questo discorso vale indipendentemente per ogni $p \leq n!$.

Teorema
Per $x \geq 2$ reale, vale $\psi(x) \geq Bx$ per una opportuna costante reale positiva B.
Dimostrazione
Dal teorema precedente, intendendo p primo, scriviamo

$$N = \frac{(2n)!}{(n!)^2} = \prod_{p \leq 2n} p^{k_p},$$

con

$$k_p = \sum_{m=1}^{\infty} \left(\left\lfloor \frac{2n}{p^m} \right\rfloor - 2\left\lfloor \frac{n}{p^m} \right\rfloor \right).$$

Possiamo notare che ogni termine della somma vale 1 oppure 0 a seconda che $\lfloor 2n/p^m \rfloor$ sia pari o dispari, in particolare vale 0 se $p^m > 2n$. Dunque

$$k_p \leq \left\lfloor \frac{\log(2n)}{\log(p)} \right\rfloor$$

e

$$\log(N) = \sum_{p \leq 2n} k_p \log(p) \leq \sum_{p \leq 2n} \left\lfloor \frac{\log(2n)}{\log(p)} \right\rfloor \log(p) = \psi(2n),$$

quest'ultima per il lemma precedente.
Tuttavia

$$N = \frac{(2n)!}{(n!)^2} = \frac{(n+1)\cdot(n+2)\cdot \ldots \cdot(2n)}{n!} = \frac{n+1}{1}\cdot\frac{n+2}{2}\cdot \ldots \cdot\frac{2n}{n} \geq 2^n,$$

dunque $\psi(2n) \geq \log(2^n) = n\log(2)$. A questo punto, per $x \geq 2$, poniamo

$$n = \left\lfloor \frac{1}{2}x \right\rfloor \geq 1$$

e abbiamo

$$\psi(x) = \psi(2n) \geq n\log(2) \geq \frac{1}{2}x\log(2),$$

che dimostra il teorema con $B = \log(2)/2$.

A questo punto, riunendo tutti i risultati precedenti, possiamo enunciare il seguente

Teorema ([10], §22.2)
Per le funzioni di Chebyshev, si ha $\psi(x) = O(x)$ e $\vartheta(x) = O(x)$.
Dimostrazione
Nei teoremi precedenti si erano mostrati i seguenti risultati

$$\vartheta(x) < Ax$$

e

$$\psi(x) \geq Bx,$$

per opportune costanti A e B precedente trovate.
Ricordando, inoltre

$$\psi(x) = \vartheta(x) + O\left(x^{\frac{1}{2}}\log^2(x)\right),$$

concludiamo

$$Bx < \psi(x) < Ax + O\left(x^{\frac{1}{2}}\log^2(x)\right),$$

cioè $\psi(x) = O(x)$. Inoltre dal teorema (§10.2.4) segue che anche $\vartheta(x) = O(x)$.

Integrale di Riemann-Stieltjes
Ci serve a questo punto introdurre brevemente l'integrale detto di Riemann-Stieltjes (o, più semplicemente, l'integrale di Stieltjes).
Inizieremo con il richiamare la definizione dell'integrale di Riemann, per poi *estenderla*.
Sia $[a,b] \subseteq \mathbb{R}$, con $a < b$. Un qualsiasi insieme (finito)

$$P = \{x_0, x_1, \ldots, x_n\}, \qquad x_i \in [a,b], \qquad a = x_0 < x_1 < \cdots < x_n = b,$$

è una partizione dell'intervallo $[a,b]$.
Consideriamo funzioni $f: [a,b] \mapsto \mathbb{R}$ continue, anche se, come sappiamo l'integrale di Riemann si può opportunamente definire per una varietà più ampia di funzioni.
Se, dunque, $f: [a,b] \mapsto \mathbb{R}$ è continua, su ogni possibile partizione P di $[a,b]$ si definiscono la *somma superiore* e la *somma inferiore* di Riemann di f relativa a P i numeri reali

$$S(f,P)=\sum_{k=1}^{n} M_k(x_k-x_{k-1}), \qquad s(f,P)=\sum_{k=1}^{n} m_k\,(x_k-x_{k-1}),$$

in cui

$$M_k=\max_{x\in[x_{k-1},x_k]} f(x), \qquad m_k=\min_{x\in[x_{k-1},x_k]} f(x).$$

Si può dimostrare che l'insieme delle somme superiori di f e quello delle somme inferiori sono separati e, se hanno un unico elemento separatore, questo è, per definizione, l'integrale di Riemann di f su $[a,b]$ che si indica con $I(f)$ o, più comunemente,

$$I(f)=\int_a^b f(x)\,dx.$$

In altre parole, detto $\mathcal{P}$ l'insieme delle partizioni di $[a,b]$, si pone

$$I^+(f)=\inf\{S(f,P): P\in\mathcal{P}\},$$
$$I^-(f)=\sup\{s(f,P): P\in\mathcal{P}\}.$$

Il primo prende il nome di integrale (di Riemann) superiore e il secondo integrale inferiore. Si dice allora che f è integrabile in $[a,b]$ secondo Riemann se e solo se

$$I^+(f)=I^-(f).$$

Questo è, appunto, $I(f)$.

Equivalentemente, f è integrabile secondo Riemann in $[a,b]$ se e solo se

$$\lim_{n\to\infty}\big(S(f,P_n)-s(f,P_n)\big)=0$$

e, in questo caso,

$$\int_a^b f(x)dx=\lim_{n\to\infty} S(f,P_n)=\lim_{n\to\infty} s(f,P_n).$$

Quest'ultima caratterizzazione ci consente di calcolare l'integrale ricorrendo alla nozione di limite invece che a quella di estremo superiore o inferiore.

Ricordiamo che l'integrale di Riemann possiede le seguenti proprietà.

(i) Per $f(x)$ e $g(x)$ funzioni integrabili in $[a,b]$ e α,β costanti, vale

$$\int_a^b [\alpha f(x)\pm\beta g(x)]dx=\int_a^b \alpha f(x)dx\ \pm\int_a^b \beta g(x)dx$$
$$=\alpha\int_a^b f(x)dx\pm\beta\int_a^b g(x)dx,$$

(proprietà che è detta anche "linearità dell'integrale".

(ii) Per $f(x)$ funzione integrabile in $[a,b]$,

$$\int_c^c f(x)=0, \qquad c\in[a,b].$$

(iii) Per $f(x)$ integrabile in $[a,b]$ e $c\in[a,b]$,

$$\int_a^b f(x)dx=\int_a^c f(x)dx+\int_c^b f(x)dx.$$

(iv) Per $f(x)$ integrabile in $[a,b]$,

$$\int_a^b f(x)dx=-\int_b^a f(x)dx.$$

L'integrale di Riemann, si può estendere nel modo che segue.

Consideriamo una funzione $f(x)$ definita come sopra e, in aggiunta, una funzione $g(x):[a,b]\mapsto\mathbb{R}$ cui si chiede semplicemente di essere continua a tratti (può avere

discontinuità e salti all'interno del suo dominio, come accade, ad esempio, per la $\psi(x)$ di Chebyshev).
Operando, dunque, nell'intervallo $[a,b]$ una partizione P, in analogia a quanto fatto in precedenza, si possono formare le due seguenti somme

$$S(f,P)=\sum_{k=1}^{n} M_k\big(g(x_k)-g(x_{k-1})\big), \qquad s(f,P)=\sum_{k=1}^{n} m_k\,\big(g(x_k)-g(x_{k-1})\big),$$

nelle quali

$$M_k=\max_{x\in[x_{k-1},x_k]} f(x), \qquad m_k=\min_{x\in[x_{k-1},x_k]} f(x).$$

Tali somme sono dette, rispettivamente, somma superiore e somma inferiore di Riemann-Stieltjes – o, più semplicemente di Stieltjes – di f su P rispetto a g.
Analogamente a prima, si può mostrare che l'insieme delle somme superiori di f e quello delle somme inferiori sono separati e, se hanno un unico elemento separatore, questo è, per definizione, l'integrale di Riemann-Stieltjes di f su $[a,b]$ che si indica con

$$I_s(f)=\int_a^b f(x)dg(x).$$

In esso, $f(x)$ è detta funzione integranda mentre $g(x)$ è la funzione integratrice.
L'integrale di Stieltjes ha le stesse proprietà di quello di Riemann (come la linearità) ma ne è un'estensione: se, infatti, ponessimo $g(x)=x$, ci riporteremmo al semplice integrale di Riemann.
Inoltre possiamo notare che, se$g(x)$ è di classe C^1 in $[a,b]$, si ha

$$\int_a^b f(x)dg(x)=\int_a^b f(x)g'(x)dx\,,$$

dove il secondo integrale è quello usuale di Riemann.
A tal proposito, basta solamente analizzare la definizione stessa dell'integrale di Stieltjes tramite le somme superiori e inferiori, osservando che

$$\big(g(x_{i+1})-g(x_i)\big)=\frac{\big(g(x_{i+1})-g(x_i)\big)}{x_{i+1}-x_i}(x_{i+1}-x_i),$$

in cui, passando al limite nella partizione, risulta

$$\frac{\big(g(x_{i+1})-g(x_i)\big)}{x_{i+1}-x_i}(x_{i+1}-x_i)\to g'(x)dx,$$

che vale solo se $g(x)$ è di classe C^1.
In questo caso, come detto, gli integrali di Riemann e di Stieltjes coincidono.

Teorema ([9], §4.3)
Vale la seguente stima asintotica:

$$\lim_{x\to+\infty}\frac{\psi(x)}{x}=1.$$

Dimostrazione
Ci serviremo della formula esplicita della funzione $\psi(x)$ vista nella sezione dei teoremi di von Mangoldt. Si era trovata, infatti, la seguente espressione valida per $x>1$

$$\psi(x)=x-\sum_{\rho}\frac{x^{\rho}}{\rho}+\sum_{n=1}^{\infty}\frac{x^{-2n}}{2n}-\frac{\zeta'(0)}{\zeta(0)}.$$

Iniziamo, dunque, con il considerare il seguente integrale

$$\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\left(-\frac{\zeta'(s)}{\zeta(s)}\right)\frac{x^{s+1}ds}{s(s+1)}, \qquad (x>1, a>1),$$

per valutarlo in due modi differenti, in analogia a quanto detto per la formula esplicita della ψ.

Il primo modo è quello di utilizzare la formula

$$-\frac{\zeta'(s)}{\zeta(s)}=\sum_{n=1}^{\infty}\Lambda(n)n^{-s},$$

analizzata nella sezione dedicata alla funzione ζ (§11.5) e richiamata in quella dedicata ai teoremi di von Mangoldt (§14.3.1). Otteniamo

$$\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\left(-\frac{\zeta'(s)}{\zeta(s)}\right)\frac{x^{s+1}ds}{s(s+1)}=\sum_{n=1}^{\infty}\Lambda(n)\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\frac{x^{s+1}ds}{s(s+1)n^s}.$$

Ora, a partire da

$$\frac{1}{s(s+1)}=\frac{1}{s}-\frac{1}{s+1},$$

otteniamo

$$\frac{x^{s+1}}{n^s s(s+1)}=\frac{x}{s}\left(\frac{x}{n}\right)^s-\frac{n}{s+1}\left(\frac{x}{n}\right)^{s+1},$$

quindi

$$\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\frac{x^{s+1}ds}{s(s+1)n^s}=\frac{x}{2\pi i}\int_{a-i\infty}^{a+i\infty}\left(\frac{x}{n}\right)^s\frac{ds}{s}-\frac{n}{2\pi i}\int_{(a+1)-i\infty}^{(a+1)+i\infty}\left(\frac{x}{n}\right)^u\frac{du}{u}=\begin{cases}x-n, & n\le x\\ 0, & n\ge x\end{cases}.$$

Nell'ultimo integrale abbiamo operato il cambio di variabile $s+1=u$ per rapportarci, in entrambi i termini, a delle formulazioni che ci consentissero di utilizzare la formula di Perron vista nella sezione dei teoremi di von Mangoldt (§14.2).

Il risultato appena mostrato ci dice che l'integrazione termine a termine nella

$$\sum_{n=1}^{\infty}\Lambda(n)\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\frac{x^{s+1}ds}{s(s+1)n^s}$$

è valida (poiché, fissato n, riguarda un numero finito di termini). Essa è uguale a

$$\sum_{n\le x}\Lambda(n)(x-n)=\int_0^x(x-t)d\psi(t),$$

nella quale si è utilizzata la proprietà dell'integrale di Riemann-Stieltjes richiamata brevemente prima di questo teorema. A questo punto dobbiamo andare nell'altra direzione dell'integrale, ci serviremo dell'espressione utilizzata per la dimostrazione della formula esplicita per la ψ (§14.3)

$$-\frac{\zeta'(s)}{\zeta(s)}=\frac{s}{s-1}-\sum_{\rho}\frac{s}{\rho(s-\rho)}+\sum_{n=1}^{\infty}\frac{s}{2n(s+2n)}-\frac{\zeta'(0)}{\zeta(0)}.$$

In analogia a quanto fatto per la dimostrazione ad opera di von Mangoldt per la formula esplicita, la calcoleremo per $s=-1$ (in precedenza per $s=0$) per poi sottrarla membro a membro alla precedente. Con dei calcoli analoghi otteniamo

$$-\frac{\zeta'(s)}{\zeta(s)}+\frac{\zeta'(-1)}{\zeta(-1)}$$
$$=\frac{s}{s-1}-\frac{-1}{-1-1}-\sum_{\rho}\left[\frac{s}{\rho(s-\rho)}-\frac{-1}{\rho(-1-\rho)}\right]$$
$$+\sum_{n=1}^{\infty}\left[\frac{s}{2n(s+2n)}-\frac{-1}{2n(-1+2n)}\right],$$

dunque

$$-\frac{\zeta'(s)}{\zeta(s)}=\frac{s+1}{2(s-1)}-\sum_{\rho}\frac{s+1}{(\rho+1)(s-\rho)}+\sum_{n=1}^{\infty}\frac{s+1}{(2n-1)(s+2n)}-\frac{\zeta'(-1)}{\zeta(-1)}.$$

Sostituendo quanto trovato nell'integrando da valutare otteniamo

$$\left(-\frac{\zeta'(s)}{\zeta(s)}\right)\frac{x^{s+1}}{s(s+1)}=-\frac{\zeta'(s)}{\zeta(s)}x^{s+1}\left(\frac{1}{s}-\frac{1}{s-1}\right)$$
$$=\frac{x^{s+1}}{s-1}-\sum_{\rho}\frac{x^{s+1}}{\rho(s-\rho)}+\sum_{n=1}^{\infty}\frac{x^{s+1}}{2n(s+2n)}-\frac{\zeta'(0)}{\zeta(0)}\frac{x^{s+1}}{s}-\frac{x^{s+1}}{2(s-1)}$$
$$+\sum_{\rho}\frac{x^{s+1}}{(\rho+1)(s-\rho)}-\sum_{n=1}^{\infty}\frac{x^{s+1}}{(2n-1)(s+2n)}+\frac{\zeta'(-1)}{\zeta(-1)}\frac{x^{s+1}}{s+1}.$$

La metodologia è proprio la stessa utilizzata nella valutazione dell'integrale per il calcolo della formula esplicita della ψ:

- si è esplicitato il termine $-\zeta'(s)/\zeta(s)$;
- si dimostra che è lecito passare dall'integrale alla somma dei termini poiché si tratta di termini convergenti (con lo stesso metodo utilizzato in precedenza);
- si sostituiscono i termini ottenuti nell'integrale rapportandosi alla formula di Perron (per i termini che hanno $(s+1)$ al denominatore, basta fare un semplice cambio di variabile $u=s+1$, sempre per utilizzare la formula di Perron).

Otteniamo, dunque,

$$\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\left(-\frac{\zeta'(s)}{\zeta(s)}\right)\frac{x^{s+1}ds}{s(s+1)}$$
$$=\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\left(-\frac{\zeta'(s)}{\zeta(s)}\right)\frac{x^{s+1}ds}{s}-\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty}\left(-\frac{\zeta'(s)}{\zeta(s)}\right)\frac{x^{s+1}ds}{s+1}$$
$$=x^2-\sum_{\rho}\frac{x^{\rho+1}}{\rho}+\sum_{n=1}^{\infty}\frac{x^{1-2n}}{2n}-\frac{\zeta'(0)}{\zeta(0)}x-\frac{x^2}{2}+\sum_{\rho}\frac{x^{\rho+1}}{\rho+1}-\sum_{n=1}^{\infty}\frac{x^{1-2n}}{2n-1}$$
$$+\frac{\zeta'(-1)}{\zeta(-1)}=\frac{x^2}{2}-\sum_{\rho}\frac{x^{\rho+1}}{\rho(\rho+1)}-\sum_{n=1}^{\infty}\frac{x^{1-2n}}{2n(2n-1)}-\frac{\zeta'(0)}{\zeta(0)}x+\frac{\zeta'(-1)}{\zeta(-1)}.$$

Abbiamo valutato l'integrale nelle due direzioni, analogamente alla dimostrazione della formula esplicita per la funzione ψ. Il risultato è, quindi, il seguente

$$\int_0^x(x-t)d\psi(t)=\frac{x^2}{2}-\sum_{\rho}\frac{x^{\rho+1}}{\rho(\rho+1)}-\sum_{n=1}^{\infty}\frac{x^{1-2n}}{2n(2n-1)}-\frac{\zeta'(0)}{\zeta(0)}x+\frac{\zeta'(-1)}{\zeta(-1)},$$

che vale per $x>1$.

A questo punto è piuttosto semplice mostrare che

$$\int_0^x (x-t)d\psi(t) \sim \frac{x^2}{2},$$

infatti

$$\int_0^x (x-t)d\psi(t) - \frac{x^2}{2} = -\sum_\rho \frac{x^{\rho+1}}{\rho(\rho+1)} - \sum_{n=1}^{\infty} \frac{x^{1-2n}}{2n(2n-1)} - \frac{\zeta'(0)}{\zeta(0)}x + \frac{\zeta'(-1)}{\zeta(-1)},$$

da cui, dividendo per $x^2/2$

$$\frac{\int_0^x (x-t)d\psi(t) - x^2/2}{x^2/2} = \frac{\int_0^x (x-t)d\psi(t)}{x^2/2} - 1$$

$$= -\sum_\rho \frac{x^{\rho-1}}{\rho(\rho+1)} - \sum_{n=1}^{\infty} \frac{x^{-1-2n}}{2n(2n-1)} - \frac{2}{x^2}\frac{\zeta'(0)}{\zeta(0)} + \frac{2}{x^2}\frac{\zeta'(-1)}{\zeta(-1)}.$$

Passando al modulo, possiamo notare che le quantità al secondo membro sono tutte infinitesime per $x \to +\infty$. L'unico problema potrebbe essere la serie

$$\sum_\rho \frac{x^{\rho-1}}{\rho(\rho+1)},$$

ma $|x^{\rho-1}| = x^r$, con $r < 0$ in quanto $Re(\rho) < 1$ (essendo ρ uno zero della ξ o, equivalentemente, uno zero non banale della ζ) dunque anche questo termine tende a zero per $x \to +\infty$.

Manca solo da dimostrare che

$$\lim_{x\to+\infty} \frac{\psi(x)}{x} = 1.$$

Sia $\varepsilon > 0$ fissato e X sufficientemente grande e tale che

$$(1-\varepsilon)\frac{x^2}{2} < \int_0^x (x-t)d\psi(t) < (1+\varepsilon)\frac{x^2}{2},$$

per ogni $x > X$. Poiché

$$\lim_{x\to+\infty} \frac{\int_0^x (x-t)d\psi(t)}{x^2/2} = 1,$$

sappiamo che tale X appena definito esiste.

Possiamo, dunque, valutare

$$\int_0^y (x-t)d\psi(t) - \int_0^x (x-t)d\psi(t),$$

con $y > x \geq X$, ottenendo

$$(1-\varepsilon)\frac{y^2}{2} - (1+\varepsilon)\frac{x^2}{2} = (1-\varepsilon)\frac{y^2-x^2}{2} - 2\varepsilon\frac{x^2}{2} < \int_0^y (x-t)d\psi(t) - \int_0^x (x-t)d\psi(t)$$

$$= \int_x^y (x-t)d\psi(t) < (1+\varepsilon)\frac{y^2}{2} - (1-\varepsilon)\frac{x^2}{2} = (1+\varepsilon)\frac{y^2-x^2}{2} - 2\varepsilon\frac{x^2}{2}.$$

Nel primo membro si è scelto un

$$(1-\varepsilon)\frac{y^2}{2} - (1+\varepsilon)\frac{x^2}{2}$$

in luogo del logico

$$(1-\varepsilon)\frac{y^2}{2} - (1-\varepsilon)\frac{x^2}{2},$$

semplicemente perché

$$(1-\varepsilon)\frac{x^2}{2} < (1+\varepsilon)\frac{x^2}{2},$$

quindi anche la nostra scelta è logica in quanto abbiamo soltanto "ampliato" l'intervallo di stima (in modo lecito) per l'integrale in modo da semplificare i calcoli successivi

$$(1-\varepsilon)\frac{y^2}{2} - (1+\varepsilon)\frac{x^2}{2} < (1-\varepsilon)\frac{y^2}{2} - (1-\varepsilon)\frac{x^2}{2}.$$

A questo punto sappiamo che la ψ di Chebyshev è una funzione crescente, dunque, utilizzando a ritroso la definizione di integrale di Stieltjes vista in precedenza, ricaviamo

$$(y-x)\psi(x) \leq \int_x^y (x-t)d\psi(t) \leq (y-x)\psi(y),$$

da cui, combinando tutte le disuguaglianze, otteniamo

$$(y-x)\psi(x) \leq \int_x^y (x-t)d\psi(t) \leq (1+\varepsilon)\frac{y^2-x^2}{2} - 2\varepsilon\frac{x^2}{2}$$

e

$$(y-x)\psi(y) \geq \int_x^y (x-t)d\psi(t) \geq (1-\varepsilon)\frac{y^2-x^2}{2} - 2\varepsilon\frac{x^2}{2}.$$

Dal fatto che $y > x$, possiamo porre $y = \beta x$ (da cui $x = y/\beta$) con $\beta > 1$, nelle due disuguaglianze

$$(\beta-1)x\psi(x) \leq (1+\varepsilon)\frac{(\beta^2-1)x^2}{2} - 2\varepsilon\frac{x^2}{2},$$

$$\left(1-\frac{1}{\beta}\right)y\psi(y) \geq (1-\varepsilon)\frac{(1-1/\beta^2)y}{2} - 2\varepsilon\frac{y^2}{2\beta^2}.$$

Dividendo i membri della prima per $(\beta-1)$ e quelli della seconda per $\left(1-\frac{1}{\beta}\right)$otteniamo

$$x\psi(x) \leq (1+\varepsilon)\frac{(\beta+1)x^2}{2} - 2\varepsilon\frac{x^2}{2(\beta-1)},$$

$$y\psi(y) \geq (1-\varepsilon)\frac{(1+1/\beta)y}{2} - 2\varepsilon\frac{y^2}{2\beta^2(1+1/\beta)}.$$

Infine, dividendo per x^2 la prima e y^2 la seconda si ha (con qualche semplice calcolo)

$$\frac{\psi(x)}{x} \leq (1+\varepsilon)\frac{(\beta+1)}{2} - \frac{\varepsilon}{\beta-1},$$

$$\frac{\psi(y)}{y} \geq (1-\varepsilon)\frac{(\beta+1)}{2} - \frac{\varepsilon}{\beta-1}.$$

Ora, nella prima disuguaglianza, la quantità a destra può essere minore di ogni numero più grande di 1 scegliendo $\beta > 1$ ma sufficientemente vicino a 1 mentre il secondo termine di questa diventa infinitesimo per ε arbitrariamente piccolo. Un discorso analogo vale per la seconda disuguaglianza nella quale si può mostrare che $\psi(y)/y$ può essere arbitrariamente maggiore di ogni numero minore di 1 per y sufficientemente grande.
Questo completa la dimostrazione del fatto che $\psi(x) \sim x$.

Teorema dei numeri primi (J. Hadamard – Ch. J. De La Vallée-Poussin)
Sia $x > 0$ e $\pi(x)$ la funzione enumerativa dei primi, allora

$$\lim_{x\to+\infty} \pi(x)\Big/\frac{x}{\log(x)} = 1.$$

Dimostrazione

Forniremo la dimostrazione tratta da ([21], §17.12); in realtà ce ne sono altre di vario tipo che svariano dall'analisi funzionale allo studio della funzione ζ. Quella qui fornita è una dimostrazione che si serve di argomenti più basilari di Analisi e Teoria Analitica dei Numeri.

Il lemma ad inizio sezione ci ha mostrato che

$$\psi(x) = \sum_{p \le x} \left\lfloor \frac{\log(x)}{\log(p)} \right\rfloor \log(p), \qquad p \text{ primo}.$$

Se $x \ge 3$ si ha

$$\left\lfloor \frac{\log(x)}{\log(p)} \right\rfloor \le \frac{\log(x)}{\log(p)}.$$

Dunque

$$\psi(x) = \sum_{p \le x} \left\lfloor \frac{\log(x)}{\log(p)} \right\rfloor \log(p) \le \sum_{p \le x} \frac{\log(x)}{\log(p)} \log(p) = \sum_{p \le x} \log(x), \qquad p \text{ primo}, \qquad x \ge 3.$$

Poiché $\pi(x)$ indica il numero dei primi non superiori ad una data quantità x, ricaviamo

$$\psi(x) \le \sum_{p \le x} \log(x) = \pi(x) \log(x).$$

Si può notare che se $3 \le y < x$ per ogni numero primo p tale che $y < p \le x$ vale

$$\frac{\log(p)}{\log(y)} > 1,$$

dunque

$$\pi(x) = \sum_{1 < p \le x} 1 = \pi(y) + \sum_{y < p \le x} 1 \le \pi(y) + \sum_{y < p \le x} \frac{\log(p)}{\log(y)}.$$

Ora moltiplichiamo la relazione ottenuta per $\log(x)/\psi(x)$

$$\begin{aligned} \frac{\pi(x)\log(x)}{\psi(x)} &\le \frac{\pi(y)\log(x)}{\psi(x)} + \frac{\log(x)}{\psi(x)} \sum_{y < p \le x} \frac{\log(p)}{\log(y)} \\ &= \frac{\pi(y)\log(x)}{\psi(x)} + \frac{\log(x)}{\psi(x)\log(y)} \sum_{y < p \le x} \log(p). \end{aligned}$$

A questo punto, ricordando che $\pi(y) \le y$ e $\sum_{y < p \le x} \log(p) < \psi(x)$, risulta

$$\frac{\pi(x)\log(x)}{\psi(x)} \le \frac{\pi(y)\log(x)}{\psi(x)} + \frac{\log(x)}{\psi(x)\log(y)} \sum_{y < p \le x} \log(p) \le \frac{y\log(x)}{\psi(x)} + \frac{\log(x)}{\log(y)}.$$

Tutte le relazioni ottenute sinora valgono per $y \ge 3$; scegliendo $y = x/\log^2(x)$ otteniamo

$$\frac{\pi(x)\log(x)}{\psi(x)} \le \frac{x\log(x)}{\log^2(x)\,\psi(x)} + \frac{\log(x)}{\log(x/log^2(x))} = \frac{x}{\log(x)\,\psi(x)} + \frac{\log(x)}{\log(x) - 2\log(\log(x))}.$$

Analizziamo separatamente l'ultimo termine, portandolo al limite per $x \to +\infty$

$$\lim_{x \to +\infty} \frac{\log(x)}{\log(x) - 2\log(log(x))} = \lim_{x \to +\infty} \frac{1}{1 - 2\log(log(x))/\log(x)} = 1,$$

nella quale si è sfruttato il limite notevole

$$\lim_{x \to +\infty} \frac{\log(log(x))}{\log(x)} = 0.$$

L'equazione, dunque, diventa

$$\frac{\pi(x)\log(x)}{\psi(x)} \leq \frac{x}{\log(x)\,\psi(x)} + 1$$

A questo punto, per il teorema precedente si ha

$$\lim_{x\to+\infty} \frac{x}{\log(x)\,\psi(x)} = 0,$$

dunque, portando al limite per $x \to +\infty$ l'equazione appena ottenuta otteniamo

$$\lim_{x\to+\infty} \frac{\pi(x)\log(x)}{\psi(x)} \leq 1.$$

Tuttavia, dalla relazione

$$\psi(x) \leq \sum_{p\leq x} \log(x) = \pi(x)\log(x)$$

otteniamo, portando anche questa al limite per $x \to +\infty$,

$$\lim_{x\to+\infty} \frac{\pi(x)\log(x)}{\psi(x)} \geq 1.$$

Possiamo, dunque, concludere

$$\lim_{x\to+\infty} \frac{\pi(x)\log(x)}{\psi(x)} = 1.$$

La dimostrazione è quasi conclusa. Ricordando nuovamente il teorema precedente otteniamo la tesi

$$1 = \lim_{x\to+\infty} \frac{\pi(x)\log(x)}{\psi(x)} = \lim_{x\to+\infty} \frac{\pi(x)\log(x)}{\psi(x)\cdot 1} = \lim_{x\to+\infty} \frac{\pi(x)\log(x)}{\psi(x)\cdot(x/\psi(x))} = \lim_{x\to+\infty} \frac{\pi(x)\log(x)}{x},$$

cioè

$$\lim_{x\to+\infty} \pi(x)\Big/\frac{x}{\log(x)} = 1.$$

Considerazioni sul teorema

Come già detto, di questo teorema esistono innumerevoli dimostrazioni e conferme derivanti da risultati successivi ad esso. La dimostrazione qui proposta, pur essendo corretta formalmente non lo è storicamente poiché si basa sul fatto che $\psi(x)\sim x$ che, in realtà, è un corollario di questo teorema. La dimostrazione originale, infatti, si basa su tecniche dell'analisi complessa e sullo studio della funzione ζ anche se in seguito si cercò di provarlo con metodi "elementari" (cioè senza gli strumenti dell'analisi complessa) oppure con i così detti teoremi Tauberiani.

- Nel 1896 il Teorema dei Numeri Primi venne dimostrato da J. Hadamard e Ch. De La Vallée-Poussin in maniera indipendente e simile servendosi del fatto che $\zeta(1+it) \neq 0$, per $t \in \mathbb{R}$ e di tecniche non elementari di analisi complessa.
- All'inizio del 1900 venne ri-dimostrato da Landau che si servì di un teorema, così detto, Tauberiano. I teoremi Tauberiani (dal matematico Tauber), fanno parte di una branca mista tra l'Analisi Complessa e la TADN e si occupano di mostrare proprietà interessanti delle funzioni integrabili. Generalmente, come corollario, forniscono delle stime per le funzioni di Chebyshev o per quella di von Mangoldt, che consentono di dimostrare in maniera "elementare" il PNT.

- Altre dimostrazioni sono quella di Hardy (1914), quella di Selberg (1953) e, ultimamente, quella di Newman (1980) basata sull'analogo teorema Tauberiano di Newman che si trova abbondantemente in rete (ad esempio, in [3]).

Ci si poteva, infatti, "accontentare" della dimostrazione iniziale del PNT, ma si è andato oltre, fornendone delle altre più semplici e comprensibili. Come detto, questo è anche un esempio di come in matematica, spesso, dopo aver ottenuto un risultato si cerca anche di semplificarlo.
Recentemente (1994), Wiles ha dimostrato l'ultimo teorema di Fermat con argomentazioni che solo una stretta élite di matematici poteva capire e confutare. Sicuramente, molte altre menti matematiche – magari fidandosi di Fermat che asserì di averne una dimostrazione "meravigliosa" (§6.1.7) – si adopereranno per darne una prova certamente elementare. In fondo Fermat visse in un'epoca nella quale, ad esempio, i numeri complessi erano ancora visti con diffidenza e mancavano due secoli alla nascita dell'Analisi Matematica.

Teorema (identità di Abel) ([3], §4.7)
Per $a(n)$ una qualsiasi funzione aritmetica, consideriamo

$$A(x) = \sum_{0 < n \leq x} a(n),$$

(da cui $A(x) = 0$ per $x < 1$). Se f è una funzione di classe C^1 nell'intervallo $[y, x]$ in cui $0 < y < x$, allora

$$\sum_{y<n\leq x} a(n)f(n) = A(x)f(x) - A(y)f(y) - \int_y^x A(t)f'(t)dt.$$

Teorema ([3], §4.3)
Per $x \geq 2$ abbiamo

$$\vartheta(x) = \pi(x)\log(x) - \int_2^x \frac{\pi(t)}{t} dt$$

e

$$\pi(x) = \frac{\vartheta(x)}{\log(x)} + \int_1^x \frac{\vartheta(t)}{t\log^2(t)} dt,$$

nelle quali $\pi(x)$ è la funzione enumerativa dei primi della quale avevamo parlato nella sezione dedicata ai numeri primi (§6.1.11).
Dimostrazione
Definiamo $a(n)$, la funzione caratteristica dei primi.
Per ogni n naturale

$$a(n) = \begin{cases} 1, & n \text{ primo}, \\ 0, & \text{altrimenti}. \end{cases}$$

Allora si ha

$$\pi(x) = \sum_{p\leq x} 1 = \sum_{1<n\leq x} a(n)$$

e

$$\vartheta(x) = \sum_{p\leq x} \log(p) = \sum_{1<n\leq x} a(n)\log(n).$$

Queste scritture si basano proprio sulla $a(n)$: nella prima e nella seconda $a(n) = 1$ solo per n primo. L'utilizzo della funzione caratteristica dei primi giustifica il passaggio dalle definizioni formali precedenti a quelle introdotte in questa dimostrazione.
Prendendo, ora, $f(x) = \log(x)$ e applicando la formula dell'identità di Abel con $y = 1$, otteniamo

$$\vartheta(x) = \sum_{1<n\leq x} a(n)\log(n) = \pi(x)\log(x) - \pi(1)\log(1) - \int_1^x \frac{\pi(t)}{t}dt,$$

che prova la prima delle due formule del teorema tenendo conto che $\pi(x) = 0$ per $x < 2$.
Ponendo, ora, $b(n) = a(n)\log(n)$ e scrivendo

$$\pi(x) = \sum_{\frac{3}{2}<n\leq x} \frac{b(n)}{\log(n)}, \qquad \vartheta(x) = \sum_{n\leq x} b(n),$$

applicando ancora l'identità di Abel con $f(x) = 1/\log(x)$ e $y = 3/2$ otteniamo

$$\pi(x) = \sum_{\frac{3}{2}<n\leq x} \frac{b(n)}{\log(n)} = \frac{\vartheta(x)}{\log(x)} - \frac{\vartheta(3/2)}{\log(3/2)} + \int_{\frac{3}{2}}^x \frac{\vartheta(t)}{t\log^2(t)}dt,$$

che prova la seconda formula poiché $\vartheta(x) = 0$ per $x < 2$.

Teorema
Le seguenti affermazioni sono logicamente equivalenti:

(i) $\lim_{x\to+\infty} \frac{\pi(x)\log(x)}{x} = 1$;

(ii) $\lim_{x\to+\infty} \frac{\vartheta(x)}{x} = 1$;

(iii) $\lim_{x\to+\infty} \frac{\psi(x)}{x} = 1$.

Se ne può trovare una dimostrazione in ([3], §4.4). In essa si applica il teorema precedente dividendo tutti i termini nella formula di $\vartheta(x)$ per x e quelli nella formula di $\pi(x)$ per $x/\log(x)$.
Da qui si può vedere proprio come il fatto che $\psi(x)\sim x$ sia equivalente al Teorema dei Numeri Primi. In altre parole da quello si può dimostrare il Teorema dei Numeri Primi (così come è stato fatto in questa sezione) e da quest'ultimo si può dimostrare che $\psi(x)\sim x$.
Nella dimostrazione originale del teorema dei numeri primi, infatti, ad opera del lavoro indipendente dei due matematici J. Hadamard e Ch. De la Vallée-Poussin, ci si serviva della funzione ζ di Riemann e del fatto che non avesse zeri per $Re(s) = 1$ (§13.2.3) per poi operare molte formule elaborate di analisi complessa non trattate in questa tesi. In seguito sono sorte altre dimostrazioni di tale teorema basate, come questa, su implicazioni successive o su risultati indipendenti che ne semplificano la trattazione.

Teorema
Sia p_n l'n −esimo numero primo. Allora le seguenti stime asintotiche sono equivalenti:

(i) $\lim_{x\to+\infty} \frac{\pi(x)\log(x)}{x} = 1$;

(ii) $\lim_{x\to+\infty} \frac{\pi(x)\log(\pi(x))}{x} = 1$;

(iii) $\lim_{n\to+\infty} \frac{p_n}{n\log(n)} = 1.$

Così come nel teorema precedente, la prima relazione è quella del Teorema dei Numeri Primi e si mostra come le altre siano equivalenti ad essa. Qui troviamo la stima asintotica dell'n −esimo numero primo.
Analogamente al teorema precedente, per la dimostrazione si rimanda a ([3], §4.4). Inoltre, in ([3], §4.6-4.9) ci sono altri risultati che derivano dal Teorema dei Numeri Primi e riguardano le somme parziali delle funzioni μ e Λ.

APPENDICE IV: FORMULA PRODOTTO DI HADAMARD PER LA ξ

In questa sezione vedremo – limitandoci agli aspetti essenziali – la dimostrazione, offerta da Hadamard nel 1893 in un suo articolo, riguardante la formula prodotto per la funzione ξ.
Come abbiamo visto nelle precedenti sezioni, la funzione ξ ha infiniti zeri. Vedremo ora una sua rappresentazione sotto forma di prodotto, che si serve appunto di questi zeri.

Introduzione alla formula prodotto

Nel suo articolo di ricerca, Riemann "dimostra" – per così dire – la formula prodotto in quattro righe molto stringate, che riportiamo brevemente qui ma che si possono trovare, per intero, nell'Appendice I.

> <<Questa funzione è finita per tutti i valori finiti di t, e permette essa stessa di essere sviluppata in potenze di tt ["tt" vale "t^2", *n.d.A.*] con una serie convergente molto rapidamente. […] Se si indicano con α tutte le radici dell'equazione $\xi(\alpha) = 0$, si può esprimere $\log \xi(t)$ come
>
> $$\sum \log\left(1 - \frac{tt}{\alpha\alpha}\right) + \log \xi(0)\,;$$
>
> e, dal fatto che la densità delle radici della quantità t cresce con t solo come $\log \frac{t}{2\pi}$, segue che questa espressione converge e diventa per un t infinito solo infinita come $t \log t$; così differisce dal $\log \xi(t)$ per una funzione di tt, che per un t finito resta continua e finita e, quando è divisa da tt, diventa infinitamente piccola per t infinito. Questa differenza è, di conseguenza, una costante, il cui valore può essere determinato ponendo $t = 0$.>>
>
> (Tratto dall'articolo di Riemann)

In queste parole Riemann enuncia – in maniera implicita e non chiarissima – una formula prodotto per la funzione ξ; da esse si evince una certa sicurezza sulla sua validità, nonostante che manchi una dimostrazione e ci si limiti a qualche considerazione sul fatto che è "ragionevole" supporla vera.
Il matematico tedesco la esprime nel modo seguente.
Egli chiama con α – ma noi abbiamo preferito la notazione ρ, largamente accettata in tutti i testi che trattano di tale argomento – le radici della funzione ξ, quindi α è tale che $\xi(\alpha) = 0$.
Successivamente trova (senza dimostrarlo)

$$\log\big(\xi(t)\big) = \sum \log\left(1 - \frac{tt}{\alpha\alpha}\right) + \log \xi(0),$$

che è la formula prodotto della ξ.
In questa scrittura ci sono due caratteristiche particolari:

- l'utilizzo di scritture come tt e $\alpha\alpha$ in luogo di t^2 e α^2;
- l'utilizzo della scrittura mediante logaritmo – ricordando che il logaritmo complesso non è iniettivo come quello reale ma, come avevamo detto in (§3.2.10), ha più rami regolari (è una funzione *a più valori*).

Riguardo a quest'ultima osservazione, prendendo la scrittura di Riemann, ed elevando ad esponenziale ambo i membri, si ottiene

$$\exp\big(\log(\xi(t))\big) = \exp\left(\sum \log\left(1 - \frac{tt}{\alpha\alpha}\right) + \log\xi(0)\right)$$
$$= \exp\left(\log\left(\prod\left(1 - \frac{tt}{\alpha\alpha}\right)\right) + \log\xi(0)\right) = \exp\left(log\left(\xi(0)\prod\left(1 - \frac{tt}{\alpha\alpha}\right)\right)\right),$$

da cui si evince

$$\xi(t) = \xi(0)\prod\left(1 - \frac{tt}{\alpha\alpha}\right).$$

Come detto nella sezione dedicata agli zeri della ζ, però, $\xi(0)$ non è $\xi(0)$ nel modo attuale in cui siamo abituati a vederlo poiché in questo caso Riemann pone

$$\xi\left(\frac{1}{2} + it\right) = \xi(s), \qquad t \in \mathbb{C},$$

dunque $\xi(0)$ corrisponde a $\xi(1/2)$ ed è per questo motivo che invece di operare una sostituzione lascia $\xi(0)$.
Non ci soffermiamo sui dettagli estetici di questa formula e della notazione utilizzata da Riemann in quanto se ne è già dibattuto nell'Appendice I.
Si può parlare a lungo dei ragionamenti di Riemann nel suo breve articolo: le discussioni possono riguardare sia la forma (alcuni di questi ragionamenti sono piuttosto "rigorosi", come nel caso delle dimostrazioni delle equazioni funzionali, altri, invece, solo "accennati" come la formula prodotto) sia la sostanza delle conclusioni del matematico tedesco.
Tuttavia in alcuni casi si tratta di affermazioni per le quali si intravede che lo stesso Riemann sembra non avere argomentazioni dettagliate ma solo giustificazioni "ragionevoli" come per la sua ipotesi (si confronti con la "lunga" discussione nell'Edwards riguardo a questo punto ([9], §1.9)).
Ricordiamo, infatti, che nel suo articolo – formula prodotto a parte – Riemann, riguardo agli zeri della ξ, afferma palesemente che

> <<[…] si trova in effetti approssimativamente questo numero di radici reali entro questi limiti, ed è molto probabile che tutte le radici sono reali. Certamente ci si augura una piccola dimostrazione qui; nel frattempo io ho temporaneamente messo da parte la ricerca per questo dopo qualche futile tentativo di sfuggita, così come sembra non necessario per il prossimo obiettivo della mia indagine.>>
>
> (Tratto dall'articolo di Riemann.)

Da queste parole si evince il fatto che Riemann avrebbe provato a dimostrare la sua ipotesi, ma dopo "qualche futile tentativo" avrebbe rinunciato poiché questo fatto non gli sembrava necessario.

E' una affermazione che, senza dubbio, contrasta con il classico *rigore matematico* al quale ci si abitua negli atenei, soprattutto nei corsi di Analisi.
Tornando al caso della formula prodotto, Riemann era a conoscenza dell'affermazione di Eulero riguardo ad un'analoga formula prodotto per la funzione seno (dimostrata successivamente ([16], §13))

$$\sin(\pi z) = \pi z \prod_{\nu=1}^{\infty} \left(1 - \frac{z^2}{\nu^2}\right).$$

In essa c'è un abuso di scrittura non indifferente.
L'indice ν, in realtà, è uno zero della funzione seno e il prodotto va inteso come "operato lungo tutti gli zeri del seno (ordinati, dunque, indicizzati mediante numeri naturali). Una scrittura formalmente (più) corretta ma non utilizzata potrebbe essere la seguente

$$\sin(\pi z) = \pi z \prod_{\nu:\sin(\nu)=0} \left(1 - \frac{z^2}{\nu^2}\right).$$

Partendo da questo questo risultato, Riemann ne ipotizza uno simile per la funzione ξ ritenendolo ragionevole per i motivi seguenti. Nell'articolo di Riemann si legge una

$$\log\big(\xi(t)\big) = \sum \log\left(1 - \frac{tt}{\alpha\alpha}\right) + \log \xi(0),$$

formalmente identica al passo intermedio trovato per giungere alla conclusione di quella del seno ([16], §13)

$$\log(\sin(\pi z)) = \log(\pi z) + \sum_{\nu=1}^{\infty} \log\left(1 - \frac{z^2}{\nu^2}\right).$$

Abbiamo visto nella sezione dedicata ai teoremi di von Mangoldt come tale formula prodotto della ξ sia un passo importante nella dimostrazione della formula esplicita per la ψ. Come abbiamo visto nell'Appendice III, tale formula esplicita è un passo fondamentale nella dimostrazione del teorema dei numeri primi.

Questioni di convergenza

Proveremo la formula prodotto della ζ seguendo il ragionamento offerto da Edwards ([9], §2) che riprende – in chiave moderna – il lavoro descritto da Hadamard nel suo articolo del 1893 "Études sur les Proprietés des Fonctions Entières et in Particulier d'une Fonction Considérée par Riemann" (in italiano "Studi sulle Proprietà delle Funzioni Intere e, in Particolare, su una Funzione Considerata da Riemann).
Supponiamo, dunque, di aver trovato la rappresentazione della ξ mediante un prodotto infinito esteso ai suoi zeri

$$\xi(s) = \xi(0) \prod_{\rho} \left(1 - \frac{s}{\rho}\right).$$

In essa, $\xi(0) = 1/2$ è una costante, dunque prendendo il logaritmo di entrambi i membri, l'interesse si sposta al prodotto in sé

$$\log(\xi(s)) = \log(\xi(0)) + \log\left(\prod_{\rho}\left(1-\frac{s}{\rho}\right)\right) = \log(\xi(0)) + \sum_{\rho}\log\left(1-\frac{s}{\rho}\right),$$

cioè al termine

$$\sum_{\rho}\log\left(1-\frac{s}{\rho}\right),$$

ed al fatto che, nel caso in cui $\xi(s) = 0$ (quindi $s = \rho$), questo termine avrà delle singolarità logaritmiche proprio perché

$$\log(\xi(s)) = \log(0) \to -\infty.$$

Riguardo questa formula, ci sono essenzialmente due problemi.

- Il primo è proprio il fatto che il logaritmo in sé è una funzione a più valori quindi non è nemmeno possibile – in linea teorica – passare attraverso di esso proprio perché occorrerebbe verificare la zona di piano complesso nella quale ci si trova. Riemann evita questo problema poiché, in questo caso, non ha molta importanza.
  Cerchiamo di formalizzare il perché di tale affermazione fissando $s \in \mathbb{C}$.
  Sappiamo che se ρ è uno zero della funzione ξ, $0 < Re(\rho) < 1$ proprio perché si era dimostrato che gli zeri della ξ si trovano lungo la così detta "striscia critica" (§13.3.1).
  Far crescere ρ in modulo equivale a dire far crescere $Im(\rho)$, poiché per $|\rho|$ sufficientemente grande l'apporto di $Re(\rho)$ nel modulo è trascurabile, essendo $0 < Re(\rho) < 1$.
  Fissato $s \in \mathbb{C}$, possiamo notare che

  $$\log\left(1-\frac{s}{\rho}\right) \to \log\left(Re\left(1-\frac{s}{\rho}\right)\right), \qquad Im(\rho) \to \infty.$$

  Questo fatto ci consente di concludere che, per ρ sufficientemente grande in modulo – o, per quanto detto, per $Im(\rho)$ sufficientemente grande – l'argomento del logaritmo è tale da restare all'interno del ramo definito dall'argomento principale (la striscia di piano di ampiezza 2π contenente l'asse reale).
  Il problema del logaritmo complesso, infatti, sta proprio nella parte immaginaria dell'argomento poiché è proprio lungo la parte immaginaria che sussistono gli infiniti valori.
  Se, infatti, fissiamo $Im(s)$, dalla formula stessa del logaritmo (§3.2.10)

  $$\log(s) = \ln|s| + i(\theta + 2k\pi),$$

  osserviamo banalmente che, per $Im(s)$ fissato, al variare di $Re(s)$ il ramo di riferimento del logaritmo complesso resta lo stesso.
  La conclusione è che la somma

  $$\sum_{\rho}\log\left(1-\frac{s}{\rho}\right)$$

  è definita poiché i multipli di $2\pi i$ che si ottengono passandola all'esponenziale sono finiti proprio perché il ramo del logaritmo tende ad essere univocamente determinato.
- Il secondo problema della somma è la sua convergenza. Passiamo ad analizzarlo nel dettaglio.

Partiamo dalla somma in questione, cioè

$$\sum_{\rho} \log\left(1-\frac{s}{\rho}\right),$$

e, in essa, vediamo cosa accade al crescere di $|\rho|$. Poiché $Re(\rho)$ è una quantità limitata (compresa tra 0 e 1), possiamo analizzare la formula al crescere di $|\rho - 1/2|$.
Dall'equazione funzionale della ξ (§13.1.1), cioè

$$\xi(s) = \xi(1-s),$$

sappiamo che ad uno zero ρ tale che $Re(\rho) < 1/2$ corrisponde automaticamente uno zero $1-\rho$ simmetrico al precedente rispetto alla linea critica $Re(s) = 1/2$. Allora

$$\sum_{\rho} \log\left(1-\frac{s}{\rho}\right) = \sum_{Im(\rho)>0} \left[\log\left(1-\frac{s}{\rho}\right) + \log\left(1-\frac{s}{1-\rho}\right)\right].$$

Questa uguaglianza è giustificata dal fatto che $\xi(s) = \xi(\bar{s})$.
Fissato uno specifico zero $\tilde{\rho}$, esso è coinvolto due volte nel secondo membro, la prima in maniera diretta, mentre la seconda come $1-\tilde{\rho}$ riferito ad un altro zero (che esiste a causa dell'equazione funzionale). Tutto questo, però, trova riscontro nel fatto che anche al primo membro troviamo $\tilde{\rho}$ conteggiato due volte: una volta in maniera ufficiale e una volta come punto coniugato del precedente (proprio perché $\xi(s) = \xi(\bar{s})$).
Per provare la convergenza della somma, dunque, basta provare la convergenza di

$$\sum_{Im(\rho)>0} \left[\log\left(1-\frac{s}{\rho}\right) + \log\left(1-\frac{s}{1-\rho}\right)\right] = \sum_{Im(\rho)>0} \log\left[\left(1-\frac{s}{\rho}\right)\left(1-\frac{s}{1-\rho}\right)\right]$$

$$= \sum_{Im(\rho)>0} \log\left[1-\frac{s(1-s)}{\rho(1-\rho)}\right].$$

Concludiamo allora che

$$\sum_{Im(\rho)>0} \log\left[1-\frac{s(1-s)}{\rho(1-\rho)}\right] = \log\left(\prod_{Im(\rho)>0}\left(1-\frac{s(1-s)}{\rho(1-\rho)}\right)\right)$$

converge se converge la seguente sommatoria

$$\sum_{Im(\rho)>0} \frac{1}{|\rho(1-\rho)|}.$$

Proveremo quanto detto servendoci dei due seguenti risultati.

Lemma ([20], §15.3)
Se $u_1, \dots, u_N$ sono numeri complessi, indicando

$$p_N = \prod_{n=1}^{N}(1+u_n), \qquad p_N^* = \prod_{n=1}^{N}(1+|u_n|),$$

allora

$$p_N^* \le e^{|u_1|+\cdots+|u_N|}$$

e inoltre

$$|p_N - 1| \le p_N^* - 1.$$

Dimostrazione
Per $x \ge 0$, la disuguaglianza $1 + x \le e^x$ è banale e segue anche dall'espansione di e^x in serie di Taylor (§1.3.2)

$$e^x = 1 + x + \frac{x^2}{2} + \cdots + \frac{x^n}{n!} + \cdots.$$

Sostituendo x con $|u_1|, \ldots, |u_N|$ per poi moltiplicare le disuguaglianze che si ottengono in tal modo a partire da quella di partenza otteniamo la prima tesi del teorema.
Per quanto riguarda la seconda, procediamo per induzione su N.
Per $N = 1$ la tesi è banale, in quanto la relazione è soddisfatta come uguaglianza

$$|1 + u_1 - 1| = |u_1| \leq |1 + |u_1| - 1| = |u_1|.$$

Supponiamo adesso la formula vera per $N = k$. La dimostreremo per $N = k + 1$ in modo da concluderne, per induzione, la sua validità.
Consideriamo, dunque, $N = k$

$$p_{k+1} - 1 = p_k(1 + u_{k+1}) - 1 = (p_k - 1)(1 + u_{k+1}) + u_{k+1}$$

quindi, tenendo conto delle proprietà del modulo,

$$|p_{k+1} - 1| \leq (p_k^* - 1)(1 + |u_{k+1}|) + |u_{k+1}| = p_{k+1}^* - 1.$$

Abbiamo, dunque, provato la formula per $N = k + 1$: per induzione concludiamo che è valida per ogni N.

Teorema ([20], §15.4)
Supponiamo che (u_n) sia una successione di funzioni complesse limitate in $\Omega \subseteq \mathbb{C}$, tali che $\sum_{n=1}^{\infty} |u_n(s)|$ converga uniformemente in Ω. Allora

$$f(s) = \prod_{n=1}^{\infty} \big(1 + u_n(s)\big)$$

converge uniformemente su Ω e, per $s_0 \in \Omega$, si ha $f(s_0) = 0$ se e solo se $u_n(s_0) = -1$ per almeno un indice n.
Inoltre, se $\{n_1, n_2, n_3, \ldots\}$ è una permutazione di $\{1,2,3, \ldots\}$, allora risulta ugualmente

$$f(s) = \prod_{k=1}^{\infty} \Big(1 + u_{n_k}(s)\Big).$$

Dimostrazione
Per ipotesi $\sum_{n=1}^{\infty} |u_n(s)|$ converge uniformemente, dunque è anch'essa limitata in Ω.
Indichiamo

$$p_N(s) = \prod_{n=1}^{N} \big(1 + u_n(s)\big),$$

cioè l'N −esimo prodotto parziale di $f(s)$.
Ora, sappiamo che le u_n sono limitate, dunque possiamo concludere, per il lemma precedente, che esiste una costante $C \in \mathbb{R}$ tale per cui $|p_N(s)| \leq C$, per ogni $s \in \Omega$ e per ogni N naturale fissato.
Scegliamo ε tale che $0 < \varepsilon < \frac{1}{2}$. Allora per definizione di convergenza uniforme della serie $\sum_{n=1}^{\infty} |u_n(s)|$, esiste un N_0 tale che

$$\sum_{n=N_0}^{\infty} |u_n(s)| < \varepsilon, \qquad (s \in \Omega).$$

Siano $\{n_1, n_2, n_3, \ldots\}$ una permutazione di $\{1,2,3, \ldots\}$ e M un intero sufficientemente grande tale che

$$\{1,2,\dots,N\} \subset \{n_1, n_2, \dots, n_M\}.$$

Se $q_M(s)$ denota l'M −esimo prodotto parziale della produttoria

$$q_M(s) = \prod_{k=1}^{M} \left(1 + u_{n_k}(s)\right),$$

allora

$$q_M - p_N = p_N \left(\frac{q_M}{p_N} - 1\right) = p_N \left[\prod_{k \neq 1,\dots,N} \left(1 + u_{n_k}\right) - 1\right].$$

In essa, tutti gli n_k sono distinti e sono più grandi di N_0. Possiamo, dunque, applicare nuovamente il lemma precedente ottenendo

$$|q_M - p_N| \leq |p_N|(e^{\varepsilon} - 1) \leq 2|p_N|\varepsilon \leq 2C\varepsilon.$$

Questa formula ha tre importanti conseguenze.

(i) Se $n_k = k$, allora $q_M = p_M$ e quest'ultima formula ci mostra che $\{p_N\}$ converge uniformemente alla funzione limite f. Questo dimostra la prima tesi del teorema poiché, per come è stata costruita, p_M converge poiché converge $\sum_{n=1}^{\infty} |u_n(s)|$ uniformemente (basta ricordare la definizione di N_0 e, dunque, di N).

(ii) Inoltre

$$\left|p_M - p_{N_0}\right| \leq 2\left|p_{N_0}\right|\varepsilon, \qquad M > N_0,$$

da cui

$$|p_M| \geq (1 - 2\varepsilon)\left|p_{N_0}\right|.$$

Inoltre

$$|f(s)| \geq (1 - 2\varepsilon)\left|p_{N_0}(s)\right|, \qquad s \in \Omega,$$

che ci mostra che $f(s) = 0$ se e solo se $p_{N_0}(s) = 0$, e cioè la seconda tesi del teorema (ricordando come è definita p_{N_0}).

(iii) q_M converge allo stesso limite di p_N, il che dimostra l'ultima tesi del teorema.

Con questi due risultati abbiamo dimostrato l'affermazione di partenza.

Ricapitolando, per dimostrare la convergenza della somma

$$\sum_{\rho} \log\left(1 - \frac{s}{\rho}\right),$$

occorre dimostrare, come detto, la convergenza di

$$\sum_{Im(\rho)>0} \log\left[1 - \frac{s(1-s)}{\rho(1-\rho)}\right] = \log\left(\prod_{Im(\rho)>0} \left(1 - \frac{s(1-s)}{\rho(1-\rho)}\right)\right)$$

e quindi, proprio per gli ultimi due risultati, la convergenza uniforme della sommatoria

$$\sum_{Im(\rho)>0} \frac{1}{|\rho(1-\rho)|}.$$

Inizieremo, dunque, trattando alcuni risultati preliminari che ci consentiranno di dimostrare la convergenza della somma in questione per poi dimostrare la formula prodotto per la ξ a partire dalla somma in questione.

Risultati intermedi

Partiremo, dunque, da alcuni teoremi che ci serviranno per dimostrare la convergenza della serie

$$\log\left(\prod_{Im(\rho)>0}\left(1-\frac{s(1-s)}{\rho(1-\rho)}\right)\right)$$

e, per quanto detto, della serie

$$\sum_{Im(\rho)>0}\frac{1}{|\rho(1-\rho)|}.$$

Teorema (Jensen) ([9], §2.2)
Sia $f(z)$ una funzione definita in $D(0,R)=\{z\in\mathbb{C}: |z|\leq R\}$ (per $R>0$) e analitica nello stesso disco. Si supponga che $f(z)$ non abbia degli zeri nella circonferenza $|z|=R$ ma ammetta zeri all'interno del disco $z_1,\dots,z_n$ (si intende che uno zero di ordine k è contato k volte in quella lista). Infine, si supponga che $f(0)\neq 0$. Allora

$$\log\left|f(0)\cdot\frac{R}{z_1}\cdot\frac{R}{z_2}\cdot\ldots\cdot\frac{R}{z_n}\right|=\frac{1}{2\pi}\int_0^{2\pi}\log\left|f\left(Re^{it}\right)\right|dt.$$

Dimostrazione
Se $f(z)$ non avesse degli zeri all'interno del disco, l'equazione sarebbe semplicemente

$$\log|f(0)|=\frac{1}{2\pi}\int_0^{2\pi}\log\left|f\left(Re^{it}\right)\right|dt,$$

che si ricava dal teorema della media di Gauss (§3.3.4) e ci dice proprio che il valore della funzione $\log|f(z)|$ al centro del disco è uguale alla media dei valori sul bordo. Dalla definizione stessa di logaritmo complesso (§3.2.10), si può osservare che $\log|f(z)|$ è la parte reale della funzione analitica $\log\big(f(z)\big)$. Riprendiamo, dunque, la formula integrale di Cauchy nel disco (§3.3.4)

$$f(z_0)\cdot Ind_\gamma(z_0)=\frac{1}{2\pi i}\int_\gamma\frac{f(z)}{z-z_0}dz,$$

nel quale z_0 è un punto qualunque interno. Nel nostro caso la funzione in questione è $\log\big(f(z)\big)$, $z_0=0$ e l'indice di avvolgimento è 1 trattandosi di una semplice circonferenza:

$$\log\big(f(0)\big)=\frac{1}{2\pi i}\int_{|z|=R}\frac{\log\big(f(z)\big)}{z}dz=\frac{1}{2\pi i}\int_0^{2\pi}\frac{\log\left(f\left(Re^{it}\right)\right)}{Re^{it}}iRe^{it}dt$$
$$=\frac{1}{2\pi}\int_0^{2\pi}\log\left(f\left(Re^{it}\right)\right)dt,$$

nel quale la funzione $\log\big(f(z)\big)$ è definita nel disco come

$$\log|f(0)|+\int_0^z\frac{f'(t)}{f(t)}dt,$$

per il teorema fondamentale del calcolo integrale.
Vogliamo applicare la formula appena vista alla funzione ausiliaria

$$F(z) = f(z) \cdot \frac{R^2 - \overline{z_1} z}{R(z - z_1)} \cdot \frac{R^2 - \overline{z_2} z}{R(z - z_2)} \cdot \ldots \cdot \frac{R^2 - \overline{z_n} z}{R(z - z_n)}.$$

Si ha

$$\log\big(F(0)\big) = \frac{1}{2\pi} \int_0^{2\pi} \log\big|F\big(Re^{it}\big)\big|\, dt,$$

poiché $F(z)$ è analitica e non ha zeri nel disco. Ma questa è la tesi del teorema poiché

$$\left|\frac{R^2 - \overline{z_j} \cdot 0}{R\big(0 - z_j\big)}\right| = \left|\frac{R}{z_j}\right|, \qquad \forall j \in \{1, \ldots, n\}.$$

Inoltre, per $|z| = R$, possiamo moltiplicare per $\bar{z}/R$ in quanto se $|z| = R$, anche $|\bar{z}| = R$, quindi il modulo non cambia. Si ottiene

$$\left|\frac{\bar{z}}{R} \cdot \frac{R^2 - \overline{z_j} z}{R\big(z - z_j\big)}\right| = \left|\frac{\bar{z}R^2 - \overline{z_j} R^2}{zR^2 - z_j R^2}\right| = \left|\frac{R^2\big(\bar{z} - \overline{z_j}\big)}{R^2\big(z - z_j\big)}\right| = 1, \qquad \forall j \in \{1, \ldots, n\}.$$

Per giungere a quella conclusione si è usata la seguente proprietà del modulo (§3.1.2)

$$z \cdot \bar{z} = |z|^2 = R^2$$

ed inoltre il fatto che il numeratore così ottenuto è il coniugato del denominatore (in modulo sono uguali). Questo risultato è l'ultimo tassello della dimostrazione poiché dimostra il motivo per cui all'interno dell'integrale nella tesi compare solamente il modulo di $f\big(Re^{it}\big)$; come abbiamo visto, gli altri termini sono tutti unitari in modulo (il loro logaritmo è nullo).

Il prossimo teorema ci fornisce una prima stima per $|\xi(s)|$.

Teorema ([9], §2.3)

Per valori di R sufficientemente grandi, la stima $|\xi(s)| \leq R^R$ è valida per $\left|s - \frac{1}{2}\right| \leq R$.

Dimostrazione

Avevamo visto (§13.1.4) l'espressione della $\xi(s)$ come serie di potenze

$$\xi(s) = \sum_{n=0}^{\infty} a_{2n} \left(s - \frac{1}{2}\right)^{2n},$$

dove a_{2n} erano gli opportuni coefficienti definiti come

$$a_{2n} = 4 \int_1^{\infty} \frac{d}{dx}\big[x^{3/2}\psi'(x)\big] x^{-1/4} \frac{\left(\frac{1}{2}\log(x)\right)^{2n}}{(2n)!} dx.$$

Il fatto che i coefficienti a_{2n} sono positivi segue immediatamente dalla definizione della funzione ψ di Jacobi (§12.3.3) e dal fatto che la x, come variabile d'integrazione, è tale che $x \geq 1$. L'unico termine di cui si poteva avere qualche dubbio è quello sotto il segno di derivata:

$$\frac{d}{dx}\left[x^{\frac{3}{2}}\psi'(x)\right] = \frac{d}{dx}\left(-\sum_{n=1}^{\infty} x^{\frac{3}{2}} n^2 \pi e^{-n^2 \pi x}\right) = \sum_{n=1}^{\infty} \left(n^4 \pi^2 x - \frac{3}{2} n^2 \pi\right) x^{\frac{1}{2}} e^{-n^2 \pi x},$$

nel quale

$$n^4 \pi^2 x - \frac{3}{2} n^2 \pi = n^2 \pi \left(n^2 \pi x - \frac{3}{2}\right) \geq 0, \qquad n \geq 1.$$

Ora, il valore più grande di $\xi(s)$ lungo il disco $\left|s-\frac{1}{2}\right| \leq R$ si ha per $s=\frac{1}{2}+R$ (in quanto i coefficienti a_n sono indipendenti da s), quindi per provare il teorema è sufficiente mostrare che

$$\xi\left(\frac{1}{2}+R\right) \leq R^R,$$

per ogni R sufficientemente grande.
Ora, ricordiamo la definizione stessa di $\xi(s)$ a partire dall'equazione funzionale della $\zeta(s)$ (§13.1.1)

$$\xi(s)=(s-1)\Gamma\left(\frac{s}{2}+1\right)\zeta(s)\pi^{-\frac{s}{2}},$$

nella quale $\zeta(s)$ decresce per $s \to +\infty$, quindi se R è fissato e se $N \in \mathbb{N}$ è scelto tale che

$$\frac{1}{2}+R \leq 2N \leq \frac{1}{2}R+2,$$

segue che

$$\xi\left(\frac{1}{2}+R\right) \leq \xi(2N)=(N!)\pi^{-N}(2N-1)\zeta(2N) \leq N^N\pi^0(2N)\zeta(2)=2\zeta(2)N^{N+1}$$

$$\leq 2\zeta(2)\left(\frac{1}{2}R+2\right)^{\frac{R}{2}+3} < R^R,$$

per R sufficientemente grande, il che prova il teorema.

Nel teorema precedente, si sono utilizzate differenti proprietà elementari delle funzioni coinvolte nell'equazione funzionale della ζ (e dunque nella definizione della ξ).

- $\zeta(s)$ è decrescente lungo direzioni parallele all'asse reale e vale $\zeta(2) \geq \zeta(2N)$, per N intero positivo fissato. Si è utilizzata per la maggiorazione successiva.
- $N! \leq N^N$, per N intero positivo fissato, anch'essa valida per la maggiorazione.
- $\pi^{-N} \leq \pi^0 = 1$, in quanto π è una costante positiva maggiore di 1.
- L'ultima minorazione stretta, cioè

  $$2\zeta(2)\left(\frac{1}{2}R+2\right)^{\frac{R}{2}+3} < R^R$$

  è valida per R opportunamente grande (ma neanche eccessivamente grande se si tiene conto che $2\zeta(2) \cong 3{,}329$ (§11.2).

Il prossimo teorema dà una stima per le radici ρ della funzione ξ.

Teorema ([9], §2.4)
Indichiamo con $N(R)$ il numero delle radici ρ della funzione $\xi(s)$ che si trovano nel disco $\left|s-\frac{1}{2}\right| \leq R$, tutte contate con la loro molteplicità. Allora

$$N(R) \leq 3R\log(R),$$

per R sufficientemente grande.
Dimostrazione
Applicando il teorema di Jensen visto ad inizio sezione alla funzione $\xi(s)$ nel disco $\left|s-\frac{1}{2}\right| \leq 2R$, si ottiene la seguente

$$\log\left(\xi\left(\frac{1}{2}\right)\right)+\sum_{|\rho-1/2|<2R}\log\left(\frac{2R}{|\rho-1/2|}\right)\leq\log((2R)^{2R})=2R\log(2R).$$

In questa sommatoria non si devono fraintendere gli indici: la notazione

$$|\rho-1/2|<2R,$$

è da intendersi come "tutte le radici della ξ che soddisfano tale condizione" che sono in un numero intero. Una notazione più corretta (ma lunga e scomoda per l'utilizzo come sommatoria) poteva essere

$$\rho\in\mathbb{C}:\xi(\rho)=0,|\rho-1/2|<2R.$$

I termini nella somma al variare di ρ sono positivi e quelli corrispondenti alle radici ρ all'interno del disco $\left|\rho-\frac{1}{2}\right|\leq R$ sono, come minimo, $\log(2)$

$$N(R)\log(2)\leq 2R\log(2R)-\frac{\log\big(\xi(1/2)\big)}{\log(2)},$$

cioè

$$N(R)\leq\frac{2}{\log(2)}R\log(2)+2R-\frac{\log\big(\xi(1/2)\big)}{\log(2)}\leq 3R\log(R),$$

per R sufficientemente grande, tale da renderla vera.

Nel caso in cui ci fossero delle radici ρ sul disco $\left|s-\frac{1}{2}\right|=2R$, il teorema di Jensen non sarebbe direttamente applicabile, ma lo si potrebbe comunque utilizzare lungo la circonferenza di raggio $R+\varepsilon$, con $\varepsilon>0$ per poi far tendere $\varepsilon\to 0$.

La convergenza e la formula prodotto

Passiamo, dunque, alla dimostrazione della convergenza della somma

$$\sum_{Im(\rho)>0}\frac{1}{|\rho(1-\rho)|}.$$

Innanzitutto, dobbiamo risistemarla a dovere

$$\sum_{Im(\rho)>0}\frac{1}{|\rho(1-\rho)|}=\sum_{Im(\rho)>0}\frac{1}{|\rho-\rho^2|}=\sum_{Im(\rho)>0}\frac{1}{|\rho^2-\rho|}=\sum_{Im(\rho)>0}\frac{1}{|\rho^2-\rho+1/4-1/4|}$$
$$=\sum_{Im(\rho)>0}\frac{1}{|(\rho-1/2)^2-1/4|}<\sum_{Im(\rho)>0}\frac{1}{|\rho-1/2|^2}.$$

Basterà, dunque, provare la convergenza della serie

$$\sum_{Im(\rho)>0}\frac{1}{|\rho-1/2|^2}$$

e il prossimo risultato servirà a questo.

Teorema

Per ogni $\varepsilon>0$ dato, la serie

$$\sum_{\rho:\,\xi(\rho)=0} \frac{1}{|\rho - 1/2|^{1+\varepsilon}}$$

converge.

Dimostrazione

Siano le radici ρ numerate $\rho_1, \rho_2, \ldots$ secondo l'ordine stabilito dalla distanza $\left|\rho - \frac{1}{2}\right|$. Sia poi $R_1, R_2, \ldots$ una sequenza di numeri reali positivi definiti implicitamente dall'equazione $4R_n \log(R_n) = n$.

Dal teorema precedente sappiamo che ci sono al massimo $3n/4$ radici ρ nel disco $\left|s - \frac{1}{2}\right| = R_n$; quindi l'$n$ −esima radice non è in questo disco, cioè,

$$\left|\rho_n - \frac{1}{2}\right| > R_n.$$

Ma allora, nella somma parziale sugli zeri

$$\sum_n \frac{1}{|\rho_n - 1/n|^{1+\varepsilon}} \leq \sum_n \frac{1}{R_n^{1+\varepsilon}} = \sum_n \frac{(4\log(R_n))^{1+\varepsilon}}{n^{1+\varepsilon}} = \sum_n \left(\frac{1}{n^{1+\varepsilon/2}} \cdot \frac{(4\log(R_n))^{1+\varepsilon}}{n^{\varepsilon/2}}\right).$$

Ora, dall'equazione $4R_n \log(R_n) = n$, ricaviamo

$$\log(n) = \log(R_n) + \log(4) + \log(log(R_n)) > \log(R_n),$$

per n opportunamente grande, dunque, portando al limite la somma parziale

$$\sum_{\rho:\,\xi(\rho)=0} \frac{1}{|\rho - 1/2|^{1+\varepsilon}} < \sum_n \left(\frac{1}{n^{1+\varepsilon/2}} \cdot \frac{n^{\varepsilon/2}}{n^{\varepsilon/2}}\right) = \sum_n \frac{1}{n^{1+\varepsilon/2}} < +\infty,$$

che dimostra il teorema.

Teorema

Sia $\varepsilon > 0$ fissato. Allora

$$Re\left(\log\left[\frac{\xi(s)}{\prod_\rho \left(1 - \frac{s - 1/2}{\rho - 1/2}\right)}\right]\right) \leq \left|s - \frac{1}{2}\right|^{1+\varepsilon},$$

per $\left|s - \frac{1}{2}\right|$ sufficientemente grande.

Dimostrazione

Fissiamo $R > 0$ e scriviamo la funzione di cui vogliamo una stima come somma di due funzioni

$$Re\left(\log\left[\frac{\xi(s)}{\prod_\rho \left(1 - \frac{s - 1/2}{\rho - 1/2}\right)}\right]\right) = u_R(s) + v_R(s),$$

in cui

$$u_R(s) = Re\left(\log\left[\frac{\xi(s)}{\prod_{|\rho - 1/2| \leq 2R} \left(1 - \frac{s - 1/2}{\rho - 1/2}\right)}\right]\right),$$

$$v_R(s) = Re\left(\log\left[\frac{1}{\prod_{|\rho-1/2|>2R}\left(1-\frac{s-1/2}{\rho-1/2}\right)}\right]\right).$$

Questi logaritmi sono definiti a meno di multipli di $2\pi i$ (a causa dei differenti rami regolari del logaritmo complesso (§3.2.10)), tuttavia le loro parti reali, come già detto, sono sempre ben definite ad eccezione dei punti $s = \rho$, per $\left|\rho - \frac{1}{2}\right| > 2R$, poiché, in essi, $u_R \to -\infty$ (si annulla $\xi(s)$ al numeratore per definizione di ρ) e $v_R \to +\infty$ (si annulla il denominatore).

E' sufficiente mostrare che, per R grande, sia u_R che v_R valgono al massimo $R^{1+\varepsilon}$ sulla circonferenza $\left|s - \frac{1}{2}\right| = R$. Da questa osservazione, infatti, segue che, se da ε passassimo a $0 < \varepsilon' < \varepsilon$, si otterrebbe

$$u_R(s) + v_R(s) \leq 2R^{1+\varepsilon'} \leq \left|s - \frac{1}{2}\right|^{1+\varepsilon},$$

nella circonferenza $\left|s - \frac{1}{2}\right| = R$ per R sufficientemente grande tale per cui $u_R \leq R^{1+\varepsilon'}$, $v_R \leq R^{1+\varepsilon'}$ e $2 \leq R^{\varepsilon-\varepsilon'}$.

Consideriamo, per prima, $u_R(s)$.

Nella circonferenza $\left|s - \frac{1}{2}\right| = 4R$, i fattori al denominatore sono tutti, al massimo, 1. Dunque

$$u_R(s) \leq Re(\log(\xi(s)) = \log|\xi(s)| \leq \log((4R)^{4R}) = 4R\log(4R) \leq R^{1+\varepsilon}$$

nella circonferenza $\left|s - \frac{1}{2}\right| = 4R$, per R sufficientemente grande tale per cui $4\log(4R) < R^{\varepsilon}$.

Ora, la funzione u_R è analitica nel disco $\left|s - \frac{1}{2}\right| = 4R$, tranne che nei punti $s = \rho$ nella zona $2R < \left|s - \frac{1}{2}\right| \leq 4R$. Ma, vicino a queste singolarità, $u_R \to -\infty$, quindi il massimo di u_R nel disco deve essere sul bordo poiché si tratta di logaritmo reale di numero reale (quindi crescente).

Ma allora il massimo di u_R nel disco, in particolare nella circonferenza $\left|s - \frac{1}{2}\right| = R$ è al massimo $R^{1+\varepsilon}$ come mostrato, in precedenza, nella circonferenza più ampia $\left|s - \frac{1}{2}\right| = 4R$ (che quindi contiene quest'ultima al suo interno).

Passiamo, ora, a considerare $v_R(s)$.

Per z complesso, nel disco $|z| \leq \frac{1}{2}$, vale la disuguaglianza

$$Re\left(\log\left(\frac{1}{1-z}\right)\right) = -Re(\log(1-z)) = Re\left(\int_0^z \frac{dt}{1-t}\right) \leq \left|\int_0^z \frac{dt}{1-t}\right| \leq |z| \max\left(\frac{1}{|1-t|}\right)$$
$$= 2|z|,$$

nella quale abbiamo utilizzato le proprietà del logaritmo ed il teorema fondamentale del calcolo integrale per la funzione $\log(1-z)$.

Allora, per $\left|s - \frac{1}{2}\right| = R$, vale la disuguaglianza

$$v_R(s) = Re\left(\log\left[\frac{1}{\prod_\rho\left(1-\frac{s-1/2}{\rho-1/2}\right)}\right]\right) \leq 2\sum_{|\rho-1/2|>2R}\frac{R^2}{|\rho-1/2|^2}$$
$$= 2\sum_{|\rho-1/2|>2R}\left(\frac{R}{|\rho-1/2|}\right)^{1-\varepsilon}\left(\frac{R}{|\rho-1/2|}\right)^{1+\varepsilon}$$
$$\leq 2\sum_{|\rho-1/2|>2R}\left(\frac{1}{2}\right)^{1-\varepsilon}\frac{R^{1+\varepsilon}}{|\rho-1/2|^{1+\varepsilon}} = 2^\varepsilon R^{1+\varepsilon}\sum_{|\rho-1/2|>2R}\frac{1}{|\rho-1/2|^{1+\varepsilon}}.$$

Per il teorema precedente, l'ultima sommatoria converge e tende a zero al crescere di R. Si ottiene, dunque $v_R(s) \leq R^{1+\varepsilon}$, per $\left|s-\frac{1}{2}\right| = R$ e R sufficientemente grande, e questo completa la dimostrazione (ricordando la definizione di u_R e v_R).

Lemma

Sia $f(s)$ una funzione analitica nel disco $|s| \leq r$ e sia $f(0) = 0$. Sia anche M il valore massimo di $Re\big(f(s)\big)$ nella circonferenza $|s| = r$. Allora per $r_1 < r$ il modulo di f sul disco interno $|s| \leq r_1$ è limitato da

$$|f(s)| \leq \frac{2r_1M}{r-r_1}, \qquad |s| \leq r_1.$$

Dimostrazione

Consideriamo la funzione ausiliaria

$$\phi(s) = \frac{f(s)}{s\big(2M-f(s)\big)}.$$

Denotiamo con $u(s)$ e $v(s)$ rispettivamente la parte reale e immaginaria di f, allora

$$|2M-u(s)| \geq M \geq u(s),$$

nella circonferenza $|s| = r$; così il modulo di ϕ in questa circonferenza è, al massimo

$$|\phi(s)| = \frac{(u^2+v^2)^{1/2}}{r[(2M-u)^2+v^2]^{1/2}} \leq \frac{(u^2+v^2)^{1/2}}{r(u^2+v^2)^{1/2}} = \frac{1}{r},$$

che implica che $|\phi(s)| \leq r^{-1}$, nel disco $|s| \leq r_1$. Tuttavia $f(s)$ si può esprimere in termini di $\phi(s)$ nel modo che segue

$$f(s) = \frac{2Ms\phi(s)}{1+s\phi(s)},$$

che mostra che, per $|s| = r_1$, il modulo di $f(s)$ è al massimo

$$|f(s)| \leq \frac{2Mr_1r^{-1}}{(1-r_1r^{-1})} = \frac{2Mr_1}{r-r_1},$$

che conclude la dimostrazione.

Teorema

Sia $f(s)$ una funzione intera che è pari, nel senso di $f(s) = f(-s)$. Supponiamo, inoltre, che questa funzione cresca più lentamente di $|s|^2$, cioè che $\forall\varepsilon > 0$ esista R tale che $Re\big(f(s)\big) < \varepsilon|s|^2$ in tutti i punti s con $|s| \geq R$. Allora la funzione è costante.

Dimostrazione

Consideriamo

$$f(s) = \sum_{n=0}^{\infty} a_n s^n,$$

cioè l'espansione di $f(s)$ in serie di potenze che esiste in quanto f è analitica.
Inoltre, dal teorema (§3.3.4) sappiamo che lo sviluppo in serie di potenze e lo sviluppo in serie di Taylor di una funzione olomorfa sono equivalenti e inoltre

$$a_n = \frac{1}{2\pi i}\int_{\gamma} \frac{f(s)ds}{s^{n+1}},$$

nel quale γ è una curva semplice chiusa contenente l'origine.
Siano ora ε e R quelli delle ipotesi del teorema e γ il disco $|s| \leq \frac{1}{2}R$. Allora, dalla formula appena vista per i coefficienti a_n, si ottiene

$$|a_n| = \left|\frac{1}{2\pi i}\int_0^{2\pi} \frac{f\left(\frac{1}{2}Re^{i\theta}\right)}{\left(\frac{1}{2}Re^{i\theta}\right)^n} i d\theta\right| \leq \frac{1}{2\pi}\int_0^{2\pi} \frac{2^n \left|f\left(\frac{1}{2}Re^{i\theta}\right)\right|}{R^n} d\theta.$$

Possiamo minorare la funzione nell'ultimo integrale sfruttando il lemma precedente, ottenendo

$$\frac{2^n}{R^n} \cdot \left|f\left(\frac{1}{2}Re^{i\theta}\right)\right| = \frac{2^n}{R^n} \cdot \frac{2(\varepsilon R^2)\left(\frac{1}{2}R\right)}{R - \left(\frac{1}{2}R\right)} = \frac{2^{n+1}\varepsilon}{R^{n-2}}.$$

Ora, per $n \geq 2$, esso è al massimo $2^{n+1}\varepsilon$ e, dal fatto che ε è arbitrario, a_n sarà nullo per $n \geq 2$.
Otteniamo, dunque, $f(s) = a_1 s$.
Tuttavia, anche $a_1 s = 0$ poiché la funzione è pari nel senso di $f(s) = f(-s)$.
Concludiamo che $f(s) \equiv a_0$, cioè è costante.

La formula prodotto

Nei precedenti paragrafi si è dimostrato che la funzione

$$F(s) = \frac{\xi(s)}{\prod_{\rho}\left[1 - \frac{s - 1/2}{\rho - 1/2}\right]}$$

è intera in quanto quoziente di due funzioni olomorfe e limitata per ogni s.
Essa, inoltre, non ha zeri, quindi il logaritmo

$$\log\big(F(s)\big) = \int_0^z \frac{F'(z)dz}{F(z)} + \log\big(F(0)\big),$$

è ben definito e $\log\big(F(0)\big)$ è determinato a meno delle costanti additive $2n\pi i$ che determinano i diversi rami regolari del logaritmo complesso.
Gli ultimi due risultati combinati insieme servono a dirci che

$$\log\big(F(s)\big) = cost = c,$$

in quanto la sua crescenza è limitata da $\left|s-\frac{1}{2}\right|^2 = O(|s|^2)$. Passando all'esponenziale, otteniamo

$$\xi(s) = c \cdot \prod_{\rho}\left(1-\frac{s-1/2}{\rho-1/2}\right),$$

nel quale c è una costante. Dividendo, ora, questo per il valore

$$\xi(0) = c \cdot \prod_{\rho}\left(1-\frac{-1/2}{\rho-1/2}\right),$$

otteniamo

$$\frac{\xi(s)}{\xi(0)} = \prod_{\rho}\left(1-\frac{s-1/2}{\rho-1/2}\right)\left(1-\frac{-1/2}{\rho-1/2}\right)^{-1} = \prod_{\rho}\left(\frac{\rho-s}{\rho-1/2}\right)\left(\frac{\rho-1/2}{\rho}\right) = \prod_{\rho}\left(1-\frac{s}{\rho}\right).$$

A questo punto, portando la costante $\xi(0)$ all'altro membro si ottiene la formula prodotto completa per la ξ, cioè

$$\xi(s) = \xi(0)\prod_{\rho}\left(1-\frac{s}{\rho}\right).$$

Questa, come già detto, dimostra che la $\xi(s)$ poteva essere scritta con un prodotto infinito convergente esteso ai suoi zeri.

Tale risultato, ora dimostrato, è alla base dei teoremi di von Mangoldt visti nella sezione dedicata e della dimostrazione del teorema dei numeri primi vista nell'Appendice III.

Si può provare, inoltre, che la formulazione di Hadamard appena ottenuta, cioè

$$\xi(s) = \xi(0)\prod_{\rho}\left(1-\frac{s}{\rho}\right),$$

è analoga a quella descritta da Riemann nel suo articolo, cioè

$$\xi(t) = \xi(0)\prod_{\alpha}\left(1-\frac{t^2}{\alpha^2}\right),$$

in cui:

- $\xi(0)$ non è $\xi(0)$ nel modo che siamo abituati ad intenderlo (§14.1.1), ma $\xi(1/2)$;
- α è il modo di indicare le radici della ξ dopo la *solita* sostituzione di Riemann

  $$\xi\left(\frac{1}{2}+it\right) = \xi(s),$$

  in cui $t \in \mathbb{C}$ [*sic*].

Per dimostrarlo basta semplicemente operare tale sostituzione nella formula prodotto per poi operare qualche calcolo ([9], §1.16).

APPENDICE V: NOTE STORICHE

Questa breve sezione è dedicata a qualche nota storica: vedremo l'ipotesi di Riemann in un contesto più ampio, dai matematici che l'hanno trattata ai risultati che le sono collegati.
Un testo divulgativo (quasi) completamente dedicato all'ipotesi di Riemann e alla sua storia piuttosto che alle caratteristiche tecniche è il volume – splendido nella lettura – di Du Satoy ([8]) dal quale sono liberamente tratti alcuni riferimenti in questa sezione.
Per gli interessati, oltre al già citato Du Satoy, rimandiamo anche alla lettura del libro di Derbyshire ([7]) che alterna capitoli tecnici a sezioni storiche molto dettagliate.

Riemann e la "sua" zeta

Come nasce l'ipotesi di Riemann?
Tralasciando la sua definizione tecnica e le varie sezioni introduttive occorse per arrivare ad essa, cerchiamo di inquadrarla nel contesto dell'articolo di Riemann. Vediamo in particolare di recuperarne la sua formulazione originaria.
Scopriamo allora che l'ipotesi di Riemann, nell'intento dell'autore, è solo quella frase – quasi in secondo piano – che compare nell'articolo a proposito della ξ:

> <<Ora, si trova in effetti approssimativamente questo numero di radici reali [della $\xi(t)$, *n.d.A.*] entro questi limiti, ed è molto probabile che tutte le radici sono reali.>>
>
> (tratto dall'articolo di Riemann)

Non si può quindi dichiarare con certezza che Riemann abbia manifestato esplicitamente nella sua nota quel grande interesse che poi altri hanno dedicato sulla sua ipotesi. Tuttavia, nel testo di Du Satoy, *l'Enigma dei Numeri Primi*, si può leggere un passaggio molto interessante a tal proposito

> <<Siegel scrisse al bibliotecario di Gottinga per chiedergli il permesso di consultare il *Nachlass* di Riemann, ovvero i suoi scritti postumi, come sono chiamati oggi. Il bibliotecario dispose che i documenti fossero spediti alla biblioteca di Francoforte dove Siegel avrebbe potuto consultarli. […] I documenti arrivarono puntualmente, ed egli si precipitò in biblioteca insieme a un collega che era in visita all'università di Francoforte. Mentre lo apriva, dal pacco fuoriuscì una gran massa di fogli zeppi di complicati calcoli numerici. Quelle pagine avrebbero smentito una volta per tutte l'immagine che di Riemann era stata data per settant'anni, ovvero quella di un matematico di intuito e concetti incapace di produrre prove solide a sostegno delle proprie idee. Indicando quella massa di calcoli, Siegel esclamò con ironia: "Eccoli qui i grandi concetti generali di Riemann!". […]

> Alcuni dei calcoli di Riemann, come quello della radice quadrata di 2 fino alla trentottesima cifra, non erano innovativi, ma altri intrigavano Siegel, che non si era mai imbattuto in qualcosa di simile. Mentre rovistava fra le pagine, quel guazzabuglio caotico di calcoli sparsi cominciò a rivelare un senso. Siegel capì che Riemann stava calcolando gli zeri del paesaggio zeta.>>
>
> (M. Du Satoy. *L'enigma dei Numeri Primi*. Cap. 7)

Il seguito di questa storia è che Siegel riuscì ad estrarre quella che è a tutt'oggi nota come formula di Riemann-Siegel (§15.2), che – sempre a detta di Du Satoy – "permetteva di calcolare le altitudini nel paesaggio zeta con estrema precisione" ([8], §7). E' opportuno precisare che Du Satoy con il termine "altitudini del paesaggio zeta" intende $|\zeta(s)|$ al variare di s complesso.
Va preso atto che Riemann, al momento del suo articolo, dice che non è suo obiettivo primario sapere se gli zeri della ξ siano o meno lungo la linea critica anche perché la sua ricerca è improntata ad un altro risultato, che è proprio il numero "dei primi minori di una certa quantità data". Tuttavia, come il passaggio del libro di Du Satoy suggerisce, è innegabile che proprio per questo motivo Riemann finì per interessarsi agli zeri della zeta e da quella sua ipotesi che aveva inizialmente abbandonato "dopo futili tentativi": ne aveva, infatti, calcolato alcuni zeri con una precisione impressionante per l'epoca (contando anche che non esistevano calcolatori).

L'ipotesi nella storia

Cerchiamo adesso di ricapitolare le motivazioni scientifiche che condussero all'ipotesi di Riemann e alla sua "inconsapevole" formulazione da parte dell'autore. La loro storia è sostanzialmente quella dei numeri primi. Possiamo fissarne l'inizio nel teorema con cui Euclide dimostra, nel libro 9 dei suoi Elementi, l'infinità dei primi: Hardy in Apologia di un matematico lo segnala come una delle pietre miliari dell'evoluzione del pensiero e della civiltà umana. Ma i numeri primi accompagnano alla loro infinità l'apparentemente incomprensibile irregolarità della loro sequenza.
Abbiamo visto del resto come gli algoritmi che dall'epoca dei Greci in poi si elaborarono per riconoscere i primi, da Eratostene a Fermat e via dicendo, si siano rivelati tutti insoddisfacenti nella pratica (fino ad AKS (§6.2.12)).
Una tappa fondamentale nell'esplorazione dei primi fu la dimostrazione che Eulero diede nuovamente della loro infinità (§10.1.6), impiegando per la prima volta strumenti di Analisi. Anche Gauss dedicò lunghi sforzi e calcoli per comprendere la sequenza dei primi e congetturò sulla loro distribuzione l'ipotesi che, ancora in termini di teoria analitica dei numeri, si formula dicendo che la funzione π che ad ogni reale positivo x associa il numero dei primi $\leq x$ si comporta asintoticamente, quando x tende a $+\infty$ come

$$\frac{x}{\log(x)},$$

o anche come $Li(x)$. Prima di Riemann, Chebyshev cercò in diversi modi di dimostrare la prima stima di Gauss. L'ipotesi di Riemann si colloca in quest'ambito: fu sostanzialmente per chiarire la congettura di Gauss che Riemann introdusse le funzioni ζ e ξ, dedicando loro il suo articolo di ricerca nel quale oltre all'ipotesi si evince la formula "esatta" per la funzione π.
E' da ricordare che a fine Ottocento J. Hadamard e Ch. de la Vallée-Poussin dimostrarono in modo indipendente la stima di Gauss per i primi. Così la congettura di Gauss prende oggi il nome di Teorema dei Numeri Primi: la dimostrazione dei due matematici prescinde ovviamente dalla dimostrazione definitiva dell'ipotesi di Riemann ma si basa sul fatto che $\zeta(1+it) \neq 0$, per $Im(s) = t \in \mathbb{R}$.

I tentativi di dimostrare l'ipotesi

Numerosi sono stati anche i tentativi di dimostrare l'ipotesi in sé.
Dopo Riemann, molti matematici si sono profusi – senza successo – in questo arduo compito: come dice anche la "zeta function song", infatti (§Appendice VI)

> <<I nomi di Landau e Bohr e Cramér,
> e Hardy e Littlewood e Tichmarsh sono qui,
> malgrado i loro sforzi, la bravura e la raffinatezza
> nel localizzare gli zeri nessuno ha avuto successo.>>
>
> (Tratto dalla traduzione della canzone della zeta.)

Hilbert, quando inserì nel 1900 l'ipotesi di Riemann nella sua famosa lista di problemi, probabilmente pensava che la sua soluzione non fosse lontana e non immaginava che essa sarebbe invece stata – dei suoi già citati 23 problemi – uno dei più resistenti, destinato a trasmettersi al secolo successivo ([8], §5). In ogni caso, la sua *profezia* di vivere abbastanza per vedere dimostrata l'ipotesi di Riemann andò in fumo.
Molte menti matematiche, dunque, si adoperarono per provare l'ipotesi o, al contrario, per confutarla. Si svilupparono infatti anche tentativi di calcolo atti a verificare se uno zero non cadesse al di fuori della linea critica in modo da smentire automaticamente l'ipotesi.
La seguente tabella ha lo scopo di far vedere, in parallelo, come si è sviluppata la ricerca in queste due direzioni. In essa riassumeremo le più importanti tappe teoriche (nella colonna relativa alla "Teoria") a fianco dei calcoli improntati ad un'eventuale smentita pratica dell'ipotesi (nella colonna relativa alla "Pratica").
In essa, l'ordine degli zeri è quello di distanza (in modulo) degli stessi dall'asse reale. Come detto non fanno testo gli zeri con parte immaginaria negativa in quanto $\zeta(s) = \zeta(\bar{s})$, per ogni $s \in \mathbb{C}$.

Teoria	**Pratica**
1859. Riemann pubblica il suo articolo di ricerca.	~ stessi anni: Riemann calcola i primi zeri non banali della ζ e trova che giacciono lungo la linea critica.
1905. Von Mangoldt prova la stima di Riemann degli zeri lungo la striscia critica.	1903. Gram calcola i primi 15 zeri della ζ notando che sono, effettivamente, lungo la linea critica.
1914. Hardy dimostra che ci sono infiniti zeri lungo la linea critica.	1914. Backlund calcola i primi 79 zeri senza trovare smentite.
1921. Hardy e Littlewood ipotizzano una stima concreta del numero degli zeri lungo la striscia critica.	1925. Hutchinson arriva a calcolare i primi 138 zeri della zeta.
1942. Selberg dimostra le congetture di Hardy e Littlewood circa la densità degli zeri lungo la striscia critica.	1953. Turing – tramite l'utilizzo "fisico" di una mdT – calcola la posizione dei primi 1141 zeri della ζ non trovando smentite.
1974. Levinson dimostra che almeno un terzo degli zeri totali giace sulla linea critica.	1977. Brent porta il numero degli zeri della ζ conosciuti a 40000000.
1987. Conrey porta la stima degli zeri non banali della ζ lungo la linea critica ad almeno il 40% del totale di tutti i possibili zeri non banali.	1986. Gli zeri conosciuti arrivano a 1500000001 grazie al lavoro di J. Van de Lune, H.J.J. te Riele e D.T. Winter

L'ipotesi, tuttavia, è ancora lontana dall'essere dimostrata.

APPENDICE VI: LA CANZONE DELLA ZETA

Questa breve sezione è dedicata e una simpatica canzone sull'ipotesi di Riemann. La si può trovare citata in vari testi dedicati all'argomento come, ad esempio, in ([7], §Appendix). La versione riportata qui è quella delle dispense del professor Zaccagnini ([28], §6.10) che la definisce "scherzosa ma istruttiva": non ci potrebbero essere termini più appropriati.
La "canzone" originale era del professore Tom Apostol – lo stesso autore di alcuni testi che si possono trovare nella bibliografia di questa tesi – e in seguito fu ampliata da Saunders Mac Lane.
Le note saranno riportate – in egual modo – sia sulla versione (originale) inglese che sulla traduzione italiana.

The Zeta Function song

(Sung to the tune of "Sweet Betsy from Pike", ([35]).)

Where are the zeros of zeta of s?
G.F.B. Riemann has made a good guess,
They're all on the critical line, said he,
And their density's one over $2\pi \log(t)$.[1]

This statement of Riemann's has been like a trigger,
And many good men, with vim and with vigor,
Have attempted to find, with mathematical rigor,
What happens to zeta as mod t get bigger.

The names of Landau and Bohr and Cramér,
And Hardy and Littlewood and Tichmarsh are there,
In spite of their efforts and skill and finesse,
In locating the zeros no one's had success.[2]

In 1914 G. H. Hardy did find,
An infinite number that lay on the line,
His theorem, however, won't rule out the case,
That there might be a zero at some other place.[3]

Let P be the function $\pi - Li$,
The order of P is not known for x high,
If square root of x times $\log(x)$ we could show,

Then Riemann's conjecture would surely be so.[4]

Related to this is another enigma,
Concerning the Lindelöf $\mu(\sigma)$,
Which measures the growth in the critical strip,
And on the number of zeros it gives us a grip.

But nobody knows how this function behaves,
Convexity tell us it can have no waves,
Lindelöf said that the shape of its graph,
Is constant when sigma is more than one half.[5]

Oh, where are the zeros of zeta of s?
We must know exactly, we cannot just guess,
In order to strengthen the prime number theorem,
The path of integration must not get too near'em.[6]

(Tom Apostol, Number Theory Conference, Caltech, Giugno 1955).

What Tom Apostol Didn't Know
André Weil has bettered old Riemann's fine guess,
By using a fancier zeta of s,
He proves that the zeros are where they should be,
Provided the characteristic is p.

There's a good moral to draw from this long tale of woe
That every young genius among you should know:
If you tackle a problem and seem to get stuck,
Just take it mod p and you'll have better luck.

(Anonimo, anche se si suppone essere Saunders Mac Lane, università di Cambridge, 1973)

What fraction of zeros on the line will be found
When mod t is kept below some given bound?
Does the fraction, whatever, stay bounded below
As the bound on mod t is permitted to grow?

The efforts of Selberg did finally banish
All fears that the fraction might possibly vanish.
It stays bounded below, which is just as it should,
But the bound he determined was not very good.

Norm Levinson managed to show, better yet,
At two-to-one odds t would be a good bet,
If over a zero you happen to trip

It would lie on the line and not just in the strip.

Lewinson tried in a classical way,
Weil brought modular means into play,
Atiyah then left and Paul Cohen quit,
So now there's no proof at all that will fit.[7]

But now we must study this matter anew,
Serre points out manifold things it makes true,
A medal might be the reward in this quest,
For Riemann's conjecture is surely the best.[8]

(Saunders Mac Lane).

Traduzione: la canzone della Funzione zeta

Dove sono gli zeri della zeta di s?
G.F.B. Riemann ha fatto una buona supposizione,
sono tutti sulla linea critica, dice lui,
e la loro densità è uno su $2\pi \log(t)$.[1]

Questa affermazione di Riemann è stata come una provocazione
e molte brave persone, con forza e vigore,
hanno provato a trovare, con matematico rigore,
cosa accade alla zeta quando il modulo di t cresce.

I nomi di Landau e Bohr e Cramér,
e Hardy e Littlewood e Tichmarsh sono qui,
malgrado i loro sforzi, la bravura e la raffinatezza
nel localizzare gli zeri nessuno ha avuto successo.[2]

Nel 1914 G. H. Hardy ha mostrato
che un numero infinito di questi giace lungo quella linea,
il suo teorema, tuttavia, non risolve il caso:
potrebbe esserci uno zero in qualche altro posto.[3]

Sia P la funzione $\pi - Li$,
l'ordine di P non si conosce per x grande,
se riuscissimo a mostrare che è la radice quadrata di x per $\log(x)$,
la congettura di Riemann sarebbe certamente vera.[4]

Legato a questo c'è un altro quesito,

riguardante la $\mu(\sigma)$ di Lindelöf
che misura la crescita lungo la striscia critica,
e sul numero degli zeri ci dà un'indicazione.

Ma nessuno sa come questa funzione si comporta,
la convessità ci dice che non ha onde,
Lindelöf disse che la forma del suo grafico,
è costante quando σ è maggiore di un mezzo.[5]

Oh, dove sono gli zeri della zeta di s?
Dobbiamo conoscerli con esattezza, non possiamo solo supporlo,
per consolidare il teorema dei numeri primi,
il cammino di integrazione non deve essere troppo vicino a questa.[6]

(Tom Apostol, conferenza sulla Teoria dei Numeri, Caltech, Giugno 1955).

Quello che Tom Apostol non sapeva
André Weil ha migliorato la buona congettura del vecchio Riemann,
usando una zeta di s più stravagante,
ha provato che gli zeri sono dove dovrebbero essere,
a condizione che la caratteristica sia p.

C'è una buona morale da cogliere da questo lungo racconto di tristezza
che ogni giovane genio in mezzo a voi dovrebbe sapere;
se affrontate un problema e sembrate bloccarvi,
basta prenderlo modulo p e avrete più fortuna.

(Anonimo, anche se si suppone essere Saunders Mac Lane, università di Cambridge, 1973)

Quale frazione degli zeri sulla linea si troverà
quando si tiene il modulo di t al di sotto di un dato limite?
Forse questa porzione, qualunque essa sia, resta inferiormente limitata
quando al limite sul modulo di t si consente di crescere?

Gli sforzi di Selberg finalmente mandarono via
tutte le paure che questa frazione potesse annullarsi.
Resta inferiormente limitata, come dovrebbe,
ma il limite che lui ha determinato non è dei migliori.

Norm Levinson riuscì a mostrare, meglio,
che in un terzo dei casi t è una buona scelta,
se su uno zero vi capita di inciampare
potrebbe giacere sulla linea e non solo sulla striscia.

Levinson ha seguito un approccio classico
mentre Weil ha coinvolto nel gioco le congruenze,
Atiyah ha lasciato perdere e Paul Cohen si è arreso,
così non c'è nessuna dimostrazione che va bene.[7]

Ma ora dobbiamo studiare la questione ancora,
Serre ha considerato delle questioni sulle varietà che la rendono vera,
una medaglia potrebbe essere la ricompensa in questo problema,
perché per la congettura di Riemann è certamente il meglio.[8]

(Saunders Mac Lane).

NOTE

1. In quattro versi: l'ipotesi di Riemann (§13.3) e la densità degli zeri (§14.1.4).
2. Si riferisce al fatto che fino ad ora nessuno è riuscito a dimostrare la verità (o la falsità) dell'ipotesi di Riemann nonostante molte – e molto grandi – menti matematiche si siano profuse in un tale sforzo.
3. Il teorema di Hardy sull'infinità degli zeri lungo la linea critica è trattato brevemente nell'Appendice II.
4. L'ipotesi di Riemann è equivalente all'affermazione (§17.2) che
$$\pi(x) = Li(x) + O\left(\sqrt{x}\log(x)\right).$$
5. Si parla dell'ipotesi di Lindelöf, da noi trattata brevemente nella sezione dedicata alle conseguenze dell'ipotesi di Riemann (§17.2).
6. Nella dimostrazione originale del Teorema dei Numeri Primi (ma anche nella dimostrazione della densità degli zeri) si prende un cammino di integrazione piuttosto ampio rispetto alla "sola" striscia critica.
7. Si parla di studi avanzati che tentano di trovare collegamenti tra l'ipotesi di Riemann e svariati ambiti della matematica (senza molto successo). Si accenna anche agli sforzi che hanno reso possibile la determinazioni di percentuali *concrete* di zeri che giacciono sulla linea critica rispetto alla totalità degli stessi.
8. Chi dovesse provare la verità o la falsità dell'ipotesi di Riemann avrebbe certamente diritto ad una medaglia Fields, il più alto riconoscimento possibile in matematica. Sempre ammesso che ci si riesca prima di aver compiuto quarant'anni ([16]).

Bibliografia

[1]. Ahlfors, L.V. *Complex Analysis, an introduction to the theory of analytic functions of one complex variable*. Second Edition, McGraw-Hill, 1966.

[2]. Apostol, T.M. *Calculus (volume 1) – One-variable calculus with an introduction to Linear Algebra*. Second Edition, John Wiley & Sons, 1967.

[3]. ——. *Introduction to Analytic Number Theory*. Springer-Verlag, New York 1976.

[4]. Baricco, A. *NOVECENTO, un monologo*. Feltrinelli, Milano 1994.

[5]. Chandrasekharian, K. *Lectures on The Riemann Zeta Function*. Bombay, 1953. Liberamente scaricabile da http://www.math.tifr.res.in/~publ/ln/tifr01.pdf.

[6]. De Agostini. *Atlante Geografico Metodico (ed. 1997-98).*

[7]. Derbyshire, J. *Prime Obsession. Bernhard Riemann and the Greatest Unsolved Problem in Mathematics*. Joseph Henry Press, Washington D.C. 2003.

[8]. Du Satoy, M. *L'enigma dei Numeri Primi*. BUR 2005.

[9]. Edwards, H. M. *Riemann's Zeta Function*. Dover Publications inc, New York, 1974.

[10]. Hardy G. H., Wright E. M. *An Introduction to the Theory of Numbers*. Fourth Edition, Oxford University Press, Londra 1960.

[11]. Leonesi S., Toffalori, C. *Numeri e Crittografia*. Springer, Milano 2006.

[12]. Magno, C. *Proprietà e applicazioni della funzione Gamma*. Liberamente scaricabile in http://www.cm-physmath.net/Gamma.pdf.

[13]. Manin, Y., Panchishkin A. *Introduction to Modern Number Theory*. Second Edition, Springer, Berlino 2005.

[14]. Marcellini P., Sbordone C. *Elementi di Analisi Matematica Uno*. Liguori Editore, Napoli 2002.

[15]. ——, Fusco N. *Elementi di Analisi Matematica Due*. Liguori Editore, Napoli 2001.

[16]. Nevanlinna R., Paatero V. *Introduction to Complex Analysis*. Addison Wesley Publishing Company. Londra 1969.

[17]. Ribenboim, P. *The New Book of Prime Number Records*. Third Edition, Springer-Verlag New York 1996.

[18]. Riemann, G. F. B. *Ueber die Anzahl del Primzahlen unter einen gegebenen Grösse*. (Trad. *Sul Numero dei Primi minori di una certa Quantità data*). Traduzione inglese dall'originale in Tedesco a cura di David R. Wilkins. Liberamente scaricabile dal sito web www.maths.tcd.ie/pub/HistMath/People/Riemann/zeta/Ezeta.pdf.

[19]. Rudin, W. *Principles of Mathematical Analysis*. Third Edition, McGraw-Hill, New York, 1976.

[20]. ——. *Real and Complex Analysis*. McGraw-Hill, New York, 1970.

[21]. Sansone, G. *Lezioni di Teoria dei Numeri (redatte per uso degli studenti)*. AA 1974-1975.

[22]. Shoup, V. *A computational Introduction to Number Theory and Algebra*. Liberamente scaricabile da http://shoup.net/ntb/.

[23]. Song Y. Yan. *Number Theory for Computing*. Second Edition, Springer, Berlino 2002.

[24]. Stein E. M., Shakarchi R. *Complex Analysis*. Princeton University Press, 2003.

[25]. Titchmarsh, E. C. *On var der Corput's method and zeta-function of Riemann (IV)*. Quarterly Journal of Mathematics, Oxford, pagg. 98-105, vol. 5, 1934.

[26]. ——. *The theory of the Riemann Zeta-Function*. Second Edition, Claredon Press, Oxford 1988.

[27]. Troianello, G. M. *Variabile Complessa*. dispensa liberamente scaricabile da internet nel seguente indirizzo http://www.mat.uniroma1.it/people/troianiello/dispense.html.

[28]. Zaccagnini, A. *Introduzione alla teoria analitica dei numeri (vers. 2008)*. Liberamente scaricabile dal sito internet dell'autore http://www.math.unipr.it/~zaccagni/home.html.

[29]. Zill D., Shanahan P. D. *A First Course in Complex Analysis, With Applications*. Jones and Bartlett Publishers, Princetown 2003.

Sitografia

[1]. http://en.wikipedia.org/wiki/Basel_problem#Euler.27s_approach.
[2]. http://en.wikipedia.org/wiki/Dirichlet_eta_function.
[3]. http://en.wikipedia.org/wiki/Gamma_function.
[4]. http://en.wikipedia.org/wiki/Lindel%C3%B6f_hypothesis.
[5]. http://en.wikipedia.org/wiki/Logarithmic_derivative.
[6]. http://en.wikipedia.org/wiki/Logarithmic_integral_function.
[7]. http://en.wikipedia.org/wiki/Mertens_conjecture.
[8]. http://en.wikipedia.org/wiki/Odlyzko%E2%80%93Sch%C3%B6nhage_algorithm.
[9]. http://en.wikipedia.org/wiki/On_the_Number_of_Primes_less_Than_a_Given_Magnitude.
[10]. http://en.wikipedia.org/wiki/Riemann%E2%80%93Siegel_formula.
[11]. http://en.wikipedia.org/wiki/Summation.
[12]. http://en.wikipedia.org/wiki/Zeta_constant.
[13]. http://it.wikipedia.org/wiki/Costante_di_Apery.
[14]. http://it.wikipedia.org/wiki/Costante_di_Eulero-Mascheroni.
[15]. http://it.wikipedia.org/wiki/Funzione_zeta_di_Riemann.
[16]. http://it.wikipedia.org/wiki/Medaglia_Fields.
[17]. http://it.wikipedia.org/wiki/Numeri_di_Bernoulli.
[18]. http://it.wikipedia.org/wiki/Numeri_primi_sexy.
[19]. http://it.wikipedia.org/wiki/Numero_di_Fermat.
[20]. http://it.wikipedia.org/wiki/O_grande.
[21]. http://it.wikipedia.org/wiki/Problemi_di_Hilbert.
[22]. http://it.wikipedia.org/wiki/Problemi_per_il_millennio.
[23]. http://it.wikipedia.org/wiki/Spazio_vettoriale#Definizione_formale.
[24]. http://mathworld.wolfram.com/HarmonicNumber.html.
[25]. http://mathworld.wolfram.com/LogarithmicIntegral.html.
[26]. http://mathworld.wolfram.com/Riemann-SiegelFunctions.html.
[27]. http://mathworld.wolfram.com/RiemannZetaFunctionZeros.html.
[28]. http://mathworld.wolfram.com/Xi-Function.html.
[29]. http://oeis.org/A059750.
[30]. Dispense scaricabili dal sito del professor A. Granwille dell'università di Montreal: http://www.dms.umontreal.ca/~andrew/.
[31]. http://www.dtc.umn.edu/~odlyzko. Nel sito del matematico Odlyzko, alla voce "some unpublished matherial", si possono trovare vari spunti interessanti riguardo la computazione degli zeri in sé (oltre a tavole numeriche). L'algoritmo per il calcolo degli zeri lungo la linea critica è descritto nel libro "The 10^{20}-th zero of the Riemann zeta function and 175 million of its neighbors".
[32]. http://www.ilmeteo.it/portale/carte-del-tempo.
[33]. http://www.wolframalpha.com.
[34]. http://www.wolframalpha.com/input/?i=plot+li%28x%29+from+x%3D0+to+6. Aprendo il collegamento, apparirà il grafico della funzione $li(x)$ così come l'ho

disegnato io: per cambiarne la scala basta aumentare o diminuire il secondo valore nella barra dell'editor di formula, laddove c'è scritto "… from 0 to 6".

[35]. http://www.youtube.com/watch?v=_8WqUenyrNE è un esempio, anche se basta inserire "zeta function song" sul motore di ricerca di youtube per trovarne svariate versioni.

[36]. http://www-dimat.unipv.it/gilardi/WEBGG/PSPDF/eulero-masch.pdf. Nella tesi, la dimostrazione è liberamente presa (o meglio, parafrasata) da questa dispensa del professor G. Gilardi dell'università di Pavia.