Dimostrazione del Teorema di Natale

Fermat's Christmas Theorem series

2025-12-25

ENG ITA

Il Teorema di Natale

Ricordiamo velocemente cosa abbiamo visto finora. Partiamo con l'enunciare il teorema di Natale in termini di congruenze.

Theorem 0.1 (Teorema di Natale)**.** *Un numero primo dispari p si può scrivere come $p = a^2 + b^2$ con $a, b \in \mathbb{Z}$ se e solo se $p \equiv_4 1$.*

Nel post precedente abbiamo dimostrato che se p si può scrivere come somma di due quadrati, allora ha resto 1 quando lo dividiamo per 4, cioè $p \equiv_4 1$. Ora vogliamo dimostrare il converso: se $p \equiv_4 1$, allora p si può scrivere come somma di due quadrati (questa è la parte difficile).

In questo post sviluppiamo tutto ciò che serve per dare una dimostrazione completa del teorema. Per prima cosa, spostiamo il problema nel contesto degli interi di Gauss, che sono un'estensione degli interi che conosciamo. Poi vedremo come le somme di due quadrati compaiono in modo naturale lì, e come riconoscere i primi che si possono scrivere come somma di due quadrati.

Interi di Gauss

Sappiamo tutti che l'equazione

$$x^2 + 1 = 0 \tag{1}$$

non ha soluzioni negli interi, perché i quadrati sono sempre non negativi, e se aggiungiamo 1 a un numero non negativo non possiamo ottenere zero. Giusto?

Bene, aggiungiamo una soluzione di questa equazione agli interi: definiamo

$$i := \sqrt{-1}$$

così otteniamo

$$i^2 + 1 = \sqrt{-1}^2 + 1 = -1 + 1 = 0$$

Ora abbiamo una soluzione di Equation 1.

Definiamo $\mathbb{Z}[i]$ come l'insieme degli interi in cui aggiungiamo questo i. I "numeri" in $\mathbb{Z}[i]$ sono della forma $a + b \cdot i$, dove a e b sono interi usuali. Per esempio, $1 + 2i$, $5 + 7i$, e $0 + i$ sono tutti interi di Gauss. I casi speciali della forma $a + 0 \cdot i$ li denotiamo semplicemente con a, mentre i numeri $0 + b \cdot i$ li denotiamo con bi. In particolare, osserva che dentro $\mathbb{Z}[i]$ ci sono tutti gli interi: infatti li recuperiamo come i numeri della forma $a + 0 \cdot i = a$.

Ma perché dovremmo lavorare negli interi di Gauss invece che negli interi classici? Lo capiremo tra poco, dal momento che l'esistenza di una soluzione della congruenza

$$x^2 + 1 \equiv_p 0$$

è strettamente legata al fatto che p sia una somma di due quadrati.

Prima di tutto: come funzionano somme e prodotti negli interi di Gauss?

È facile: basta trattare i come una variabile formale, e ricordare che $i^2 = -1$. Lo spiego meglio:

Supponiamo di voler sommare $3 + 2i$ con $5 + i$. Allora facciamo così:

$$(3 + 2i) + (5 + i) = (3 + 5) + (2 + 1)i = 8 + 3i$$

cioè sommiamo separatamente le parti senza i e quelle con i. Altri esempi:

$$(-2 + 4i) + (7 - 3i) = 5 + i, \qquad (3 + 6i) - i = 3 + 7i, \qquad 3 + (-1 + i) = 2 + i.$$

Analogamente per i prodotti: trattiamo i come una variabile formale:

$$(3 + 2i) \cdot (5 + i) = 3 \cdot 5 + 3 \cdot i + 2i \cdot 5 + 2i \cdot i = 15 + 3i + 10i + 2i^2 = 15 + 2i^2 + 13i$$

Ora ricordiamo che $i^2 = -1$, quindi $2i^2 = -2$, e dunque

$$15 + 2i^2 + 13i = 15 - 2 + 13i = 13 + 13i$$

Altri esempi:

$$(-2 + 4i)(7 - 3i) = -2 + 34i, \qquad (3 + 6i) \cdot (-i) = 6 - 3i, \qquad 3 \cdot (-1 + i) = -3 + 3i$$

Ora definiamo quanto segue:

Definizione: Sia $a + bi$ un intero di Gauss. Definiamo il suo coniugato come $a - bi$.

Example 0.1.

- Il coniugato di $3+4i$ è $3-4i$.
- Il coniugato di $1-i$ è $1+i$.
- Il coniugato di 4 è 4.
- Il coniugato di $-6i$ è $6i$.

Ora osserva questo fatto utile. Se moltiplichi un intero di Gauss $a+bi$ per il suo coniugato $a-bi$ ottieni

$$(a+bi)(a-bi) = a^2 - abi + abi - b^2i^2 = a^2 - b^2i^2$$

e siccome $i^2 = -1$ otteniamo

$$(a+bi)(a-bi) = a^2 + b^2$$

È bello, no? Il nostro obiettivo è capire quando un primo è una somma di due quadrati, e ora abbiamo un modo semplice per produrre somme di due quadrati.

Cos'è un primo negli interi di Gauss?

Se ti ricordi, nel secondo post abbiamo definito un primo p come un numero tale che, ogni volta che p divide un prodotto $a \cdot b$, allora p divide a oppure p divide b. Ora facciamo lo stesso negli interi di Gauss. Prima definiamo cosa significa dividere qualcosa negli interi di Gauss.

Definizione: Sia $a+bi$ un intero di Gauss. Diciamo che $a+bi$ divide $c+di$ se possiamo scrivere

$$c+di = (a+bi)\cdot(e+fi)$$

per qualche intero di Gauss $e+fi$.

Definizione: Sia $a+bi$ un intero di Gauss. Diciamo che $a+bi$ è primo se ogni volta che $a+bi$ divide un prodotto $(c+di)(e+fi)$, allora $a+bi$ divide $c+di$ oppure $a+bi$ divide $e+fi$.

Non dimostreremo il seguente teorema, la cui dimostrazione è in un certo senso simile a quella fatta per gli interi.

Theorem 0.1 (Fattorizzazione unica)**.** *Ogni intero di Gauss $a+bi$ si può scrivere come prodotto di primi di Gauss, cioè*

$$a+bi = (c_1+d_1i)\cdots(c_k+d_ki)$$

per qualche intero positivo k. Questa scrittura è unica a meno di moltiplicare per $-1, i,$ e $-i$.

Chiariamo questa unicità, che è un po' strana. Quello che stiamo dicendo è che se consideri un intero di Gauss come $1+3i$, lo puoi fattorizzare come

$$1+3i = (1+i)(2+i)$$

(puoi verificare che $1+i$ e $2+i$ sono primi usando la definizione, ma può essere un po' noioso). Ora osserva che $i \cdot (-i) = -i^2 = 1$. Quindi possiamo scrivere

$$1+3i = 1 \cdot (1+i)(2+i) = (i \cdot (-i))(1+i)(2+i) = i \cdot ((-i) \cdot (1+i))(2+i) = i \cdot (1-i)(2+i)$$

questa sembra un'altra fattorizzazione di $1+3i$. Tuttavia differisce dalla precedente solo perché abbiamo moltiplicato per $1 = i \cdot (-i)$. Quindi diciamo che, a meno di moltiplicare per questi numeri speciali (se ci pensi puoi fare lo stesso anche con $1 = (-1) \cdot (-1)$ e $1 = (-i) \cdot i$), la fattorizzazione è unica.

Dimostrazione del teorema (finalmente!)

Ok, ci siamo quasi. Ora possiamo provare a capire come funzionerà la dimostrazione.

Sia p un primo negli interi. Uno schema della dimostrazione è mostrare le seguenti affermazioni:

- se $x^2 + 1 \equiv_p 0$ ha una soluzione allora $p + 0i$ non è primo negli interi di Gauss;
- se $p + 0i$ non è primo negli interi di Gauss, allora p si può scrivere come somma di due quadrati;
- se p è somma di due quadrati allora $p \equiv_4 1$ (lo abbiamo già dimostrato l'ultima volta!);
- se $p \equiv_4 1$ allora $x^2 + 1 \equiv_p 0$ ha una soluzione.

Queste implicazioni ci dicono che se sappiamo che una delle affermazioni sopra è vera, allora anche le altre tre lo sono. In altre parole, le affermazioni sopra stanno dicendo la stessa cosa: sono equivalenti, perché si implicano a vicenda (nota che la prima ipotesi coincide con l'ultima conclusione).

Relazione tra $x^2 + 1 \equiv_p 0$ e il teorema di Natale

Il seguente lemma è cruciale: grazie a questo lemma potremo dimostrare il teorema di Natale. Dà l'interpretazione giusta di cosa significa che un primo p sia una somma di due quadrati.

Lemma 0.1. *Sia p un numero primo. Se la congruenza*

$$x^2 + 1 \equiv_p 0$$

ha una soluzione, allora $p + 0i$ non è primo negli interi di Gauss.

Il significato di "$x^2 + 1 \equiv_p 0$ ha una soluzione" è che esiste un intero $x \in \mathbb{Z}$ tale che $x^2 + 1 \equiv_p 0$, oppure equivalentemente che p divide $x^2 + 1$. Un altro modo di dirlo è

$$x^2 \equiv_p -1$$

quindi stiamo dicendo che -1 è un quadrato modulo p.

Proof. Supponiamo che esista un $x \in \mathbb{Z}$ tale che

$$x^2 + 1 \equiv_p 0$$

Questo significa che p divide $x^2 + 1$. Ma osserva che

$$x^2 + 1 = (x + i)(x - i)$$

negli interi di Gauss. Vogliamo dire che $p + 0i$ non è primo negli interi di Gauss.

Supponiamo $p + 0i$ primo negli interi di Gauss. Siccome $p + 0i$ divide il prodotto $(x + i)(x - i)$, allora dovrebbe dividere uno dei due fattori. Ma è impossibile. Infatti, se $p + 0i$ divide $x + i$, allora possiamo scrivere $x + i = (p + 0i)(a + bi) = pa + pbi$. Confrontando i coefficienti di i sui due lati otteniamo $1 = pb$, impossibile. Lo stesso argomento vale per $x - i$.

Quindi $p + 0i$ non è primo negli interi di Gauss. □

Primi negli interi che non sono primi negli interi di Gauss

Lemma 0.2. *Sia p un numero primo. Se $p + 0i$ non è primo in $\mathbb{Z}[i]$, allora possiamo scrivere p come somma di due quadrati.*

Proof. Se $p + 0i$ non è primo, allora si può scrivere come prodotto di due fattori:

$$p + 0i = (a + bi)(c + di)$$

dove né $a + bi$ né $c + di$ sono uguali a $1, -1, i, -i$.

Ora moltiplichiamo entrambi i lati per i loro coniugati. A sinistra otteniamo

$$(p + 0i)(p - 0i) = p^2.$$

A destra otteniamo

$$(a + bi)(a - bi)(c + di)(c - di) = (a^2 + b^2)(c^2 + d^2).$$

Quindi

$$p^2 = (a^2 + b^2)(c^2 + d^2).$$

Abbiamo $a^2 + b^2 > 1$ e $c^2 + d^2 > 1$. Ma p è primo in $\mathbb{Z}$, quindi l'unico modo in cui due interi maggiori di 1 possono moltiplicarsi per dare p^2 è che

$$a^2 + b^2 = p, \qquad c^2 + d^2 = p.$$

In particolare,

$$p = a^2 + b^2,$$

cioè p è una somma di due quadrati. □

Teorema di Wilson e soluzione di $x^2 + 1 \equiv_p 0$

Useremo il seguente teorema per dimostrare ciò che ci serve.

Theorem 0.1 (Teorema di Wilson)**.** *Se p è un numero primo, allora*

$$(p-1) \cdot (p-2) \cdot (p-3) \cdots 2 \cdot 1 \equiv_p p - 1.$$

cioè moltiplicando tutti i numeri da $p-1$ fino a 1 otteniamo $p-1$ modulo p.

Diamo un'idea di come dimostrare questo teorema.

Idea. Ricorda che un intero x è invertibile modulo p se e solo se $\text{MCD}(x, p) = 1$. Siccome p è primo, non ha fattori in comune con nessuno dei numeri $(p-1), (p-2), \ldots, 2, 1$. Questo significa che ogni fattore a sinistra è invertibile modulo p.

Ora, se elenchi tutti i residui invertibili modulo p, ottieni esattamente $1, 2, \ldots, p-1$. Quindi, a sinistra, ogni termine compare insieme al suo inverso, tranne i termini che sono il proprio inverso.

Un elemento a è il proprio inverso se e solo se $a^2 \equiv_p 1$. Modulo un primo, questo significa $a \equiv_p 1$ oppure $a \equiv_p -1$. Quindi tutto si cancella a coppie, e rimane solo $-1 \equiv_p p - 1$. □

Lemma 0.3. *Se p è un primo tale che $p \equiv_4 1$ allora $x^2 + 1 \equiv_p 0$ ha una soluzione.*

Proof. Dato che $p \equiv_4 1$, significa che $p-1$ è divisibile per 4, cioè $p - 1 = 4k$ per qualche intero k. Consideriamo

$$(p-1) \cdot (p-2) \cdot (p-3) \cdots 3 \cdot 2 \cdot 1.$$

Ora raggruppiamo i fattori a coppie così:

$$(p-1) \cdot 1 \cdot (p-2) \cdot 2 \cdots (p-2k) \cdot (2k).$$

Siccome $p - j \equiv_p -j$, ogni coppia dà

$$(p-j) \cdot j \equiv_p (-j) \cdot j = -j^2.$$

Quindi

$$(p-1) \cdot (p-2) \cdots 2 \cdot 1 \equiv_p (-1)^2 \cdot (-2)^2 \cdots (-(2k))^2 = (1 \cdot 2 \cdot 3 \cdots \cdot (2k-1) \cdot 2k)^2.$$

(Qui i segni meno spariscono perché stiamo elevando al quadrato.)

Ma per il teorema di Wilson Theorem 0.1 sappiamo anche che il lato sinistro è congruo a $p-1$ modulo p. E siccome $p - 1 \equiv_p -1$, otteniamo

$$-1 \equiv_p (1 \cdot 2 \cdots (2k-1) \cdot 2k)^2$$

Definiamo

$$x := 1 \cdot 2 \cdots (2k-1) \cdot 2k$$

Allora otteniamo

$$x^2 + 1 \equiv_p 0.$$

Quindi abbiamo trovato una soluzione. □

Il teorema è dimostrato, cos'altro possiamo dire?

Il teorema è finalmente dimostrato, ma non smettiamo di fare domande. Ne vengono fuori due.

Quanti primi si possono scrivere come somma di due quadrati? Equivalentemente, quanti primi hanno resto 1 quando li dividiamo per 4?

Probabilmente non sorprende che ci aspettiamo che ci siano infiniti primi scrivibili come somma di due quadrati. Questa affermazione è lasciata come esercizio guidato alla fine del post.

Un fatto interessante, ma non banale, è capire anche la densità di questi primi. Cioè: se consideriamo tutti i primi fino a un intero N, quanti di questi ci aspettiamo che siano somme di due quadrati? La risposta è che ci aspettiamo circa la metà. Ecco una tabella che può convincerti:

N	primi $\leq N$	primi che sono $a^2 + b^2$	proporzione
10^2	25	12	0.48
10^3	168	81	0.4821 ...
10^6	78498	39176	0.49907 ...
10^9	50847534	25423492	0.499994 ...

Questo si può dimostrare, ma è ben oltre i nostri interessi qui.

Un'altra domanda naturale è

Dato un intero n, quando si può scrivere come somma di due quadrati?

Quindi stiamo generalizzando dal caso dei primi al caso di un intero qualunque. Questo non è difficile da dimostrare conoscendo il teorema di Natale, quindi lo lasciamo come esercizio guidato alla fine.

Conclusione

Vorrei sottolineare che il punto cruciale del nostro ragionamento è stato collegare il problema di scrivere p come somma di due quadrati con la soluzione della congruenza $x^2 + 1$. Infatti, se ci pensi, gli interi di Gauss sono definiti aggiungendo agli interi una soluzione di $x^2 + 1 = 0$. Non è una coincidenza.

Esistono teoremi più generali che spiegano precisamente come la fattorizzazione di un (ideale) primo in un'estensione degli interi (ottenuta aggiungendo qualcosa di simile alle soluzioni di un polinomio) sia legata alla fattorizzazione (nel nostro caso, alle soluzioni) del polinomio modulo quel primo.

Nel nostro caso sapevamo che $x^2 + 1 \equiv_p 0$ ha una soluzione, e questo equivale a dire che esiste un intero u tale che $u^2 \equiv_p -1$. Allora modulo p abbiamo la fattorizzazione

$$x^2 + 1 \equiv_p (x + u)(x - u),$$

perché

$$(x + u)(x - u) = x^2 - u^2 \equiv_p x^2 + 1.$$

Questo ci dice esattamente che il primo p si può scrivere come prodotto di due primi negli interi di Gauss (osserva la simmetria: il polinomio si spezza in due fattori esattamente come il primo). Il teorema che spiega questo comportamento è noto come teorema di Dedekind-Kummer.

Esercizi guidati

Esercizio 1: infiniti primi che sono somme di due quadrati

Esercizio 1

Dimostra che esistono infiniti primi che si possono scrivere come somma di due quadrati.

Suggerimento 1

Un primo è una somma di due quadrati se e solo se -1 è un quadrato mod p. Supponi che ci siano solo finitamenti molti primi tali che -1 sia un quadrato mod p. Cosa puoi dire?

Soluzione al Suggerimento 1

Dire "-1 è un quadrato mod p" significa: esiste un intero n tale che

$$n^2 \equiv_p -1,$$

cioè $p \mid (n^2+1)$. Quindi ogni primo per cui -1 è un quadrato mod p divide almeno un numero della forma n^2+1.

Suggerimento 2

Sia k il prodotto di tutti i primi p tali che -1 sia un quadrato mod p. Prova a imitare la dimostrazione di Euclide che esistono infiniti primi.

Suggerimento 3

Considera l'intero k^2+1. Quali primi lo possono dividere?

Soluzione al Suggerimento 3 (e fine dell'esercizio)

Sia q un primo che divide k^2+1. Allora

$$k^2 \equiv_q -1,$$

quindi -1 è un quadrato mod q. Per definizione di k, questo forza $q \mid k$.

Ma allora q divide sia k sia k^2+1, quindi divide anche la loro differenza $(k^2+1)-k^2=1$, impossibile. Contraddizione.

Quindi l'ipotesi "solo finitamenti molti primi con -1 quadrato mod p" è falsa, ed esistono infiniti primi di questo tipo. Per il teorema di Natale, questi sono esattamente i primi che sono somme di due quadrati.

Esercizio 2: quali interi sono somme di due quadrati?

Esercizio 2

Quali interi $n \geq 1$ si possono scrivere come $n = a^2 + b^2$?

Suggerimento 1

Cosa succede se n è il prodotto di due primi che si possono scrivere come somma di due quadrati?

Soluzione al Suggerimento 1

Se $p = a^2 + b^2$ e $q = c^2 + d^2$, allora anche pq è una somma di due quadrati. Infatti una somma di due quadrati si può scrivere come un intero di Gauss per il suo coniugato:

$$a^2 + b^2 = (a + bi)(a - bi), \qquad c^2 + d^2 = (c + di)(c - di).$$

Quindi

$$(a^2 + b^2)(c^2 + d^2) = (a + bi)(a - bi)(c + di)(c - di).$$

Il fattore $(a + bi)(c + di)$ è un intero di Gauss, e $(a - bi)(c - di)$ è il suo coniugato. Quindi il loro prodotto è una somma di due quadrati.

Suggerimento 2

Considera il numero $5 \cdot 7^2$. Si può scrivere come somma di due quadrati?

Soluzione al Suggerimento 2

Sì:

$$5 \cdot 7^2 = 245 = 14^2 + 7^2.$$

(Quindi il primo $7 \equiv_4 3$ non è "vietato", ma sembra contare l'esponente.)

Suggerimento 3

Nel suggerimento precedente, $5 \equiv_4 1$ e $7 \equiv_4 3$. Che teorema ti aspetti?

Soluzione al Suggerimento 3 (enunciato)

Teorema (somma di due quadrati). Un intero positivo n si può scrivere come $n = a^2 + b^2$ se e solo se, nella fattorizzazione prima di n, ogni primo $q \equiv_4 3$ compare con esponente **pari**.

Dimostrazione completa (guidata dai suggerimenti)

Passo 1: i primi $q \equiv_4 3$ devono comparire un numero pari di volte.

Dimostriamo prima il seguente lemma:

Lemma 0.1. *Sia q un primo con $q \equiv_4 3$. Se $q \mid (a^2 + b^2)$, allora $q \mid a$ e $q \mid b$.*

Proof. Supponiamo $q \mid (a^2 + b^2)$. Se $q \nmid b$, allora b ha un inverso u modulo q, e

$$(a \cdot u)^2 \equiv_q -1,$$

quindi -1 è un quadrato mod q. Ma per il teorema di Natale, questo forzerebbe $q \equiv_4 1$, contraddizione. Quindi $q \mid b$. Allora $q \mid a^2$ e quindi $q \mid a$. □

Ora supponi $n = a^2 + b^2$. Prendi un primo $q \equiv_4 3$ che divide n. Per il lemma, q divide sia a sia b, quindi q^2 divide $a^2 + b^2 = n$. Questo significa che nella fattorizzazione prima di n l'esponente di q non può essere dispari: deve essere pari.

Passo 2: il converso.

Ora supponi che n abbia la proprietà che ogni primo $q \equiv_4 3$ compaia con esponente pari. Allora possiamo pensare n come prodotto di tre tipi di fattori:

- una potenza di 2,
- primi $p \equiv_4 1$ (anche ripetuti),
- e quadrati di primi $q \equiv_4 3$ (anche ripetuti).

  Ognuno di questi fattori è una somma di due quadrati:

- $2 = 1^2 + 1^2$,
- se $p \equiv_4 1$ allora p è una somma di due quadrati per il teorema di Natale,
- e se hai un quadrato q^{2m} allora $q^{2m} = (q^m)^2 + 0^2$.

  Infine, per il Suggerimento 1, un prodotto di somme di due quadrati è ancora una somma di due quadrati. Quindi n è una somma di due quadrati.