3 Interi somma di due quadrati

Uno dei primi problemi che Fermat prese in esame fu quello della rappresentazione dei numeri naturali come somma di due quadrati di interi, cercando di chiarire alcuni passaggi dell'*Arithmetica* di Diofanto che trattavano di tale argomento. Anche F. Viète, C. Bachet e A. Girard, suoi contemporanei, si occuparono di tale questione anche se, come Fermat, non diedero dimostrazioni complete dei risultati a cui pervennero.

Diamo alcuni esempi delle questioni esaminate da Diofanto:

- determinare $a, b, c \in \mathbb{Z}$ tali che $a + b = 1$ e che $a + c$ e $b + c$ sono entrambi quadrati; in tale situazione, dunque, $2c + 1$ è la somma di due quadrati;
- 15 non è mai somma di due quadrati;
- determinare $a, b, c, d \in \mathbb{Z}$ in modo tale che, posto $q := (a + b + c + d)^2$, $q \pm a$, $q \pm b$, $q \pm c$, $q \pm d$ sono tutti quadrati.

Diofanto osserva che quest'ultima questione può essere facilmente risolta se si trovano quattro diversi triangoli pitagorici aventi la stessa ipotenusa o, ciò che è equivalente, se si trova un quadrato che può essere espresso in quattro modi diversi come somma di due quadrati. A questo proposito, Fermat enunciò che, se p è un primo del tipo $4k + 1$, allora p è sempre uguale al quadrato della lunghezza dell'ipotenusa di un unico triangolo a cateti interi (cfr. i successivi Teorema 3.4 e Corollario 3.6).

Il problema che vogliamo esaminare in questo paragrafo è quello di trovare gli interi n per i quali l'equazione diofantea in due indeterminate

$$X^2 + Y^2 = n$$

è risolubile. I principali risultati di questo paragrafo furono dimostrati completamente per la prima volta da Euler.

Il seguente enunciato era verosimilmente già noto a Diofanto e veniva comunque riportato da Leonardo da Pisa (detto Fibonacci) nel suo celebre *Liber Abaci* del 1202.

Proposizione 3.1. *Siano $n, m \in \mathbb{N}^+$. Se n e m possono essere scritti come somma di due quadrati di interi, allora anche nm può essere scritto come somma di due quadrati di interi.*

Dimostrazione. Semplice conseguenza della seguente identità:

$$(a^2+b^2)(c^2+d^2) = (ac \mp bd)^2 + (ad \pm bc)^2 \ .$$

□

Osservazione 3.2. La dimostrazione data da Euler nel 1770 del risultato precedente (con la scelta "superiore" dei segni) si basa sul fatto che tale relazione esprime, con linguaggio moderno, la proprietà moltiplicativa della norma nell'anello degli interi di Gauss. (Cioè, se $\alpha := a+bi, \beta := c+di \in \mathbb{Z}[i]$, allora $\mathrm{N}(\alpha) = a^2+b^2$, $\mathrm{N}(\beta) = c^2+d^2$ e $\mathrm{N}(\alpha\beta) = \mathrm{N}(\alpha)\mathrm{N}(\beta) = (a^2+b^2)(c^2+d^2) = (ac-bd)^2+(ad+bc)^2$.)

Proposizione 3.3. *Sia p un numero primo. Se $p \equiv 3 \pmod 4$, allora p non può essere scritto come somma di due quadrati di interi.*

Dimostrazione. Se per assurdo $x^2+y^2 = p$, allora $x^2+y^2 \equiv 3 \pmod 4$ e ciò è assurdo in quanto, come ben noto, x^2 e y^2 sono congrui a 0 oppure a 1 (mod 4).

□

Il risultato seguente fu comunicato in una lettera da Fermat a Mersenne nel 1640. Tuttavia una proprietà analoga era stata enunciata precedentemente da Girard. La prima dimostrazione completa di tale risultato fu data da Euler nel 1747.

Teorema 3.4. (P. Fermat, 1640) *Un primo p è esprimibile come somma di due quadrati di interi se, e soltanto se, $p = 2$ oppure $p \equiv 1 \pmod 4$.*

Dimostrazione. Se p è un primo esprimibile come somma di due quadrati, allora, per la Proposizione 3.3, $p \not\equiv 3 \pmod 4$. Viceversa, se $p = 2$ allora $2 = 1^2+1^2$; non è dunque restrittivo supporre che $p \equiv 1 \pmod 4$. In tal caso, sappiamo che la congruenza $X^2 \equiv -1 \pmod p$ ammette soluzioni (cfr. Proposizione I.6.6 (h)). Sia a una soluzione di tale congruenza, sia cioè $a^2+1 = hp$, con $h \geq 1$. Se ne deduce subito che $\mathrm{MCD}(a,p) = 1$. Consideriamo la congruenza lineare:

$$aX \equiv Y \pmod p$$

ed utilizziamo il seguente:

Lemma 3.5. (A. Thue, 1902) *Sia p un primo e a un intero tale che* $\mathrm{MCD}(a,p)=1$. *Allora la congruenza*

$$aX \equiv Y \pmod{p}$$

ammette una soluzione x_0, y_0, con $x_0, y_0 \in \mathbb{Z}$ tali che

$$0 \lneqq |x_0| \lneqq \sqrt{p} \qquad 0 \lneqq |y_0| \lneqq \sqrt{p} \ .$$

Tale lemma è una semplice applicazione del seguente:

Principio di Dirichlet (noto anche come ***Principio delle gabbie di piccioni o delle caselle postali***). *Se un insieme di n elementi (piccioni) deve essere ripartito in m sottoinsiemi (gabbie), cioè se ogni elemento (piccione) deve essere assegnato ad un sottoinsieme (gabbia), e se $n \gneqq m \geq 1$, allora un sottoinsieme (almeno) contiene più di un elemento (cioè, in una stessa gabbia debbono trovarsi almeno due piccioni).*

Pur essendo intuitivamente ovvio, il Principio di Dirichlet è un "teorema" e come tale necessita di una dimostrazione. Sia, per assurdo, falso e sia n il minimo intero positivo per cui è falso. Necessariamente risulta $n \geq 2$. Sia quindi S un insieme con n elementi e supponiamo che questi siano ripartiti in m sottoinsiemi, $S_1, \ldots, S_m$, $n \gneqq m \geq 1$, in modo tale che nessun sottoinsieme contenga due o più elementi di S. Ovviamente $m > 1$. Se a è un elemento di S, allora a appartiene esattamente ad un sottoinsieme, diciamo S_1 (per fissare le idee). Consideriamo allora l'insieme $S' := S \setminus \{a\}$. Gli elementi di S' rimangono ripartiti nei sottoinsiemi $S_2, \ldots, S_m$, in modo tale che nessun sottoinsieme contiene due o più elementi di S'. Ma S' ha $n-1$ elementi e questo contraddice la minimalità di n.

Dimostrazione del Lemma 3.5. Si ponga $k := [\sqrt{p}] + 1$ e si consideri l'insieme $S := \{ax - y : 0 \leq x \leq k-1,\ 0 \leq y \leq k-1\}$.

È ovvio che $\#S \leq k^2$.

Se $\#S < k^2$, siamo arrivati perché $ax_1 - y_1 = ax_2 - y_2$, per una qualche scelta delle coppie $(x_1, y_1) \neq (x_2, y_2)$; quindi basta porre $x_0 := x_1 - x_2$, $y_0 := y_1 - y_2$.

Se $\#S = k^2$, essendo $k^2 \gneqq p$, per il Principio di Dirichlet, devono esistere due elementi $ax_1 - y_1$, $ax_2 - y_2 \in S$ in modo tale che

$$ax_1 - y_1 \equiv ax_2 - y_2 \pmod{\mathrm{p}}$$

con $x_1 \neq x_2$ oppure $y_1 \neq y_2$. Quindi

$$a(x_1 - x_2) \equiv y_1 - y_2 \pmod{\mathrm{p}} \ .$$

La conclusione si ottiene ponendo $x_0 := x_1 - x_2$, $y_0 := y_1 - y_2$. Infatti, essendo $\mathrm{MCD}(a,p) = 1$ e non potendo essere x_0 e y_0 entrambi nulli, si ricava facilmente che $x_0 \neq 0$ e $y_0 \neq 0$.

□

Fine della dimostrazione del Teorema 3.4. Sia (x_0, y_0), con $x_0, y_0 \in \mathbb{Z}$, una soluzione della congruenza $aX \equiv Y \pmod{p}$ tale che $0 \lneqq |x_0|, |y_0| \lneqq \sqrt{p}$. Dunque $-x_0^2 \equiv a^2 x_0^2 \equiv y_0^2 \pmod{p}$, ovvero $x_0^2 + y_0^2 \equiv 0 \pmod{p}$. Pertanto, esiste un intero $t \geq 1$, in modo tale che $x_0^2 + y_0^2 = tp$. Dal momento che $|x_0|$, $|y_0| \lneqq \sqrt{p}$, allora $tp = x_0^2 + y_0^2 \lneqq 2p$, dunque necessariamente $t = 1$.

□

Corollario 3.6. (L. Euler, 1754) *Ogni primo p, tale che $p \equiv 1 \pmod{4}$, può essere scritto come somma di due quadrati di interi positivi in modo unico, a meno dell'ordine degli addendi.*

Dimostrazione. Supponiamo che $p = a^2 + b^2 = c^2 + d^2$, dove a, b, c, d sono interi positivi. Allora, si ha

$$p^2 = (a^2 + b^2)(c^2 + d^2) = (ad + bc)^2 + (ac - bd)^2 \ .$$

D'altra parte:

$$p(d^2 - b^2) = (a^2 + b^2)d^2 - (c^2 + d^2)b^2 = a^2 d^2 - c^2 b^2 \equiv 0 \pmod{p} \ ,$$

quindi $(ad - cb)(ad + cb) \equiv 0 \pmod{p}$. Dunque,

$$ad \equiv cb \pmod{p} \quad \text{oppure} \quad ad \equiv -cb \pmod{p} \ .$$

Essendo $0 \lneqq a, b, c, d \lneqq \sqrt{p}$, allora si ha:

$$ad - cb = 0 \quad \text{oppure} \quad ad + cb = p \ .$$

Se $p = ad + cb$, allora essendo:

$$p^2 = (ad + bc)^2 + (ac - bd)^2 = p^2 + (ac - bd)^2 \ ,$$

si ricava che $ac - bd = 0$. Dunque, si ha che:

$$ad - cb = 0 \quad \text{oppure} \quad ac - bd = 0 \ .$$

Supponiamo, per esempio, che $ad = bc$, allora $a \mid bc$ e $\mathrm{MCD}(a,b) = 1$, dunque $a \mid c$. Quindi $ah = c$, per qualche intero $h \geq 1$, e $d = bh$, donde

$p = c^2 + d^2 = h^2(a^2 + b^2) = h^2 p$. Ne segue che $h = 1$, cioè $a = c$ e $b = d$. Un ragionamento analogo, nel caso in cui $ac = bd$, permette di concludere. □

Prima di enunciare il teorema fondamentale di questo paragrafo osserviamo che ogni intero positivo n può essere rappresentato in una ed una sola maniera nella forma $n = \ell^2 m$ con ℓ, m interi positivi ed m privo di fattori quadratici (semplice conseguenza del Teorema Fondamentale dell'Aritmetica).

Teorema 3.7. *Sia $n = \ell^2 m > 0$ un intero positivo ed m un intero positivo privo di fattori quadratici. Allora n può essere rappresentato come somma di due quadrati di interi se, e soltanto se, per ogni primo dispari p tale che $p \mid m$, risulta $p \equiv 1 \pmod 4$.*

Dimostrazione. Sia $m = p_1 \dots p_r$ la fattorizzazione in primi (non necessariamente distinti) di m. Supponiamo che per ogni primo p_i, $1 \leq i \leq r$, risulti $p_i \equiv 1 \pmod 4$ oppure $p_i = 2$. Per induzione su r, applicando il Teorema 3.4 e la Proposizione 3.1, si ha che m è una somma di due quadrati e quindi, poiché $\ell^2 = \ell^2 + 0^2$, anche n è una somma di due quadrati.

Viceversa, se $n = a^2 + b^2 = \ell^2 m$ con $a, b \geq 0$, allora, posto $d := \mathrm{MCD}(a, b)$ e $a = da'$, $b = db'$, dove $\mathrm{MCD}(a', b') = 1$, allora

$$n = d^2(a'^2 + b'^2) = \ell^2 m \ .$$

Sia p un fattore primo dispari di m (senza perdere di generalità, si può supporre che m ne possieda uno, altrimenti la conclusione sarebbe immediata). Quindi,

$$a'^2 + b'^2 = (\ell/d)^2 m = tp$$

per qualche $t > 0$. Inoltre, risulta $\mathrm{MCD}(a', p) = 1$, infatti se $p \mid a'$, allora $p \mid b'$ e $\mathrm{MCD}(a', b') \neq 1$. Dunque, deve esistere un intero α' in modo tale che $a'\alpha' \equiv 1 \pmod p$. Essendo $a'^2 + b'^2 \equiv 0 \pmod p$, moltiplicando per α'^2 si ottiene $1 + (\alpha' b')^2 \equiv 0 \pmod p$, cioè $\left(\frac{-1}{p}\right) = 1$ e quindi, (cfr. Proposizione I.6.6 (h)), si ricava che $p \equiv 1 \pmod 4$.

□

Osservazione 3.8. A. Von Wijngarden nel 1950 aveva prodotto una tavola delle rappresentazioni di $n = x^2 + y^2$ con $x, y \in \mathbb{N}, 0 \leq x \leq y$, per $n \leq 10.000$. Con l'uso di mezzi di calcolo sempre più potenti è possibile ora trovare rappresentazioni di n come somma di due quadrati per valori di n molto grandi.

Corollario 3.9. (P. Fermat, 1640) *Sia $n > 0$ un intero tale che $n = x^2 + y^2$ con $x, y \in \mathbb{N}$ e sia p un numero primo dispari.*

(a) *Se* $\mathrm{MCD}(x, y) = 1$ *e* $p \mid n$, *allora risulta* $p \equiv 1 \pmod 4$ *e la congruenza* $X^2 \equiv -1 \pmod n$ *è risolubile.*

(b) *Se* $p \mid n$ *e* $p \equiv 3 \pmod 4$, *allora una potenza di* p *con esponente pari deve dividere* n , *più precisamente:*

$$n = p^{2k} n' = (p^k x')^2 + (p^k y')^2$$

dove k, n', x', y' *sono interi opportuni, con* $k, n' > 0$.

Dimostrazione. (a) Ragionando come nella dimostrazione del Teorema 3.7, poiché $p \mid n$ e $\mathrm{MCD}(x, y) = 1$, si ha che $p \nmid x$. Quindi, esiste un intero x^* tale che $xx^* \equiv 1 \pmod p$ ed, essendo $x^2 + y^2 \equiv 0 \pmod p$, si ottiene $1 + (x^* y)^2 \equiv 0 \pmod p$, da cui $p \equiv 1 \pmod 4$ (cfr. Proposizione I.6.6 (h)) e $1 + (x^* y)^2 \equiv 0 \pmod n$ (cfr. Teorema I.6.36).

(b) Supponiamo che $p^h \mid n$ e $p^{h+1} \nmid n$. Poiché $p \equiv 3 \pmod 4$ si ha, per il punto (a), che $\mathrm{MCD}(x, y) = d \gneqq 1$. Ponendo $x = dx_1$, $y = dy_1$ e $N = n/d^2$, allora si ha $N = x_1^2 + y_1^2$ con $\mathrm{MCD}(x_1, y_1) = 1$. Se $p^k \mid d$ e $p^{k+1} \nmid d$, allora $p^{h-2k} \mid N$. Ciò è assurdo (cfr. (a)) a meno che $h = 2k$.

□

3 Esercizi e complementi

3.1. Siano $a, b \in \mathbb{N}^+$ tali che $\mathrm{MCD}(a,b) = 1$. Mostrare che, se a non è somma di due quadrati di interi, allora ab non è somma di due quadrati di interi.

[*Suggerimento.* Si noti che $ab = p_1^{e_1} \dots p_t^{e_t} q_1^{f_1} \dots q_r^{f_r}$, dove $a = \prod_{i=1}^{t} p_i^{e_i}$, $b = \prod_{j=1}^{r} q_j^{f_j}$ sono le fattorizzazioni in primi distinti di a e b ed inoltre risulta $p_i \neq q_j$ per ogni i e per ogni j, perché $\mathrm{MCD}(a,b) = 1$.

Si osservi che l'ipotesi $\mathrm{MCD}(a,b) = 1$ è essenziale. Infatti se $a = 3$ e $b = 6$, allora a non è somma di due quadrati di interi, però $a \cdot b = 18 = 3^2 + 3^2$.]

3.2. Sia $a \in \mathbb{N}^+$. Mostrare che se a non è somma di due quadrati di interi, allora a non può essere neanche somma di due quadrati di numeri razionali.

[*Suggerimento.* Se, per assurdo, $a = \left(\frac{x}{y}\right)^2 + \left(\frac{z}{w}\right)^2$ con $x, y, z, w \in \mathbb{Z}$, $yw \neq 0$, allora $a(yw)^2 = (xw)^2 + (yz)^2$. Utilizzando il Teorema 3.7 si giunge facilmente ad una contraddizione.]

3.3. Mostrare che:

(a) Un numero razionale $\alpha = a/b$, con $a, b \in \mathbb{Z}$, $b \neq 0$, è somma di due quadrati di numeri razionali se, e soltanto se, ab è somma di due quadrati di interi.

(b) Un numero razionale $\alpha = a/b$, con $a, b \in \mathbb{Z}$, $b \neq 0$, $\mathrm{MCD}(a,b) = 1$, è somma di due quadrati di numeri razionali se, e soltanto se, a e b sono entrambi somma di due quadrati di interi.

(c) Se un numero razionale è somma di due quadrati di numeri razionali, allora esso ha una infinità di rappresentazioni distinte come somma di due quadrati di numeri razionali (positivi).

[*Suggerimento.* (a) Se $\frac{a}{b} = \left(\frac{c}{d}\right)^2 + \left(\frac{e}{f}\right)^2$ allora $ab(df)^2 = (bcf)^2 + (bde)^2$ e quindi per il Teorema 3.7, ab deve essere somma di due quadrati. Se $ab = x^2 + y^2$ allora $\frac{a}{b} = \left(\frac{x}{b}\right)^2 + \left(\frac{y}{b}\right)^2$.

(b) Si noti che, per il Teorema 3.7, se $\mathrm{MCD}(a,b) = 1$ allora ab è somma di due quadrati se e soltanto se a e b sono entrambi somma di due quadrati.

(c) Se $\gamma = \alpha^2 + \beta^2$ con $\alpha, \beta \in \mathbb{Q}$ e $\alpha, \beta > 0$ allora si può verificare che

$$\gamma = \alpha_k^2 + \beta_k^2$$

con $\alpha_k := \left(\frac{(k^2-1)\alpha - 2k\beta}{k^2+1}\right)^2$, $\beta_k := \left(\frac{(k^2-1)\beta - 2k\alpha}{k^2+1}\right)^2$, con $k \geq 1$ intero.]

3.4. Mostrare che, preso comunque $r \in \mathbb{N}^+$:

(a) Il numero naturale $n = a^2$, dove

$$a = (3^2+1)(4^2+1)\dots((r+2)^2+1)$$

è somma di due quadrati di interi non negativi in (almeno) r maniere distinte. Più precisamente:

$$n = a_k^2 + b_k^2 \ , \quad \text{per } k = 3, 4, \ldots, r+2 \ ,$$

dove

$$a_k = (k^2-1)a/(k^2+1) \ , \quad b_k = 2ka/(k^2+1) \ .$$

(b) Esistono sempre almeno r triangoli pitagorici distinti, aventi la stessa ipotenusa.

[*Suggerimento*. (a) Verifica diretta; (b) segue da (a)].

3.5. Verificare che, se $n = x^2 + y^2$, con $x, y \in \mathbb{Z}$, allora:

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

[La verifica è immediata.]

3.6. (**P. Fermat**, 1640). Mostrare che se p è un primo dispari del tipo x^2+2, allora p non può dividere un intero n del tipo y^2-2.

[*Suggerimento*. Si noti che $x^2+y^2 = (x^2+2)+(y^2-2)$, quindi se $p = x^2+2$ divide y^2-2, allora $p \mid (x^2+y^2)$. Essendo $p = x^2+2$ dispari deve essere $x^2 \equiv 1 \pmod 4$, quindi $p \equiv 3 \pmod 4$.]

3.7. Per ogni $n \geq 1$, si ponga:

$$r_2(n) := \#\{(x,y) \in \mathbb{Z} \times \mathbb{Z} : \ x^2+y^2 = n\} \ .$$

Da un punto di vista "geometrico" la funzione $r_2(n)$ esprime il numero dei punti del piano a coordinate intere che giacciono sulla circonferenza di equazione

$$X^2 + Y^2 = n \ , \quad n > 0 \ . \tag{3.7.1}$$

Evidentemente, se un punto (x,y) giace sulla circonferenza di equazione (3.7.1), allora anche i punti $(-x,y)$, $(x,-y)$, $(-x,-y)$ vi giacciono. Inoltre, tali punti sono tutti distinti se, e soltanto se, $xy \neq 0$. Se invece (x,y) è una soluzione intera di (3.7.1), con $xy = 0$, allora, e soltanto allora, n è un quadrato. In tal caso, $n = a^2$ con $a \in \mathbb{Z}$ e quindi $(x,y) \in \{(a,0),(0,a),(-a,0),(0,-a)\}$.

Per ogni $n \in \mathbb{N}^+$, si ponga:

$${}_2r(n) := \#\{(x,y) \in \mathbb{Z} \times \mathbb{Z} : x^2+y^2 = n \ , \quad 0 \leq x \ , \quad 0 \lneqq y\} \ .$$

Da un punto di vista geometrico, ${}_2r(n)$ esprime il numero dei punti a coordinate intere *del primo quadrante* che giacciono sulla circonferenza di equazione (3.7.1). Evidentemente:

$$r_2(n) = 4 \cdot {}_2r(n) \ .$$

Mostrare che:

(a) $r_2(1) = 4$, $r_2(2) = 4$, $r_2(3) = 0$, $r_2(4) = 4$, $r_2(5) = 8$, $r_2(6) = 0$, $r_2(7) = 0$, $r_2(8) = 4$, $r_2(9) = 4$, $r_2(10) = 8$.

(b) $r_2(n) \leq 4\sqrt{n}$.

(c) Non esiste una costante K in modo tale che, per ogni $n \in \mathbb{N}^+$, risulti $r_2(n) \leq K$.

(d) Per ogni $n \in \mathbb{N}^+$, $r_2(n) = r_2(2n)$.

(e) Se p è un numero primo, allora:

$$r_2(p) = \begin{cases} 4 & \text{se } p = 2 \\ 8 & \text{se } p \equiv 1 \pmod 4 \\ 0 & \text{se } p \equiv 3 \pmod 4 \end{cases} .$$

(f) ${}_2r(n)$ è una funzione (aritmetica) moltiplicativa, mentre $r_2(n)$ è una funzione aritmetica che non è in generale moltiplicativa.

[*Suggerimento*. (a) La verifica è diretta. (b) Basta mostrare che ${}_2r(n) \leq \sqrt{n}$. Ciò è semplice conseguenza del fatto che $x^2 + y^2 = n \Rightarrow |x| \leq \sqrt{n}$ (e, poi, $|y| = \sqrt{n - x^2}$). L'affermazione (c) discende dal precedente Esercizio 3.4 (a). L'enunciato (d) è una conseguenza del precedente Esercizio 3.5. (e) Dal punto (a) si ha che $r_2(2) = 4$ e dal Teorema 3.7 discende che $r_2(p) = 0$ se $p \equiv 3 \pmod 4$. Se $p \equiv 1 \pmod 4$, si ha che ${}_2r(p) = 2$. Infatti, per il Corollario 3.6, le due soluzioni si determinano scambiando l'ordine delle coordinate. (f) $r_2(n)$ non è moltiplicativa perché $r_2(10) = 8 \neq r_2(2)r_2(5) = 4 \cdot 8$ (cfr. (a)). L'affermazione che ${}_2r(n)$ è una funzione moltiplicativa discende immediatamente dalla Proposizione 3.1 .]

3.8. (a) Mostrare che se $n \in \mathbb{N}$, $n \geq 2$, è tale che la congruenza:

$$X^2 \equiv -1 \pmod n \tag{3.8.1}$$

è risolubile, allora ogni soluzione a di (3.8.1) determina un'unica coppia di interi (x, y) in modo tale che:

(i) $n = x^2 + y^2$, $x > 0$, $y > 0$, $\mathrm{MCD}(x, y) = 1$,

(ii) $ax \equiv y \pmod n$.

Viceversa, data una coppia di interi (x, y), soddisfacente alla condizione (i), allora questa determina, tramite la congruenza (ii), un'unica soluzione a della congruenza (3.8.1).

(b) Per ogni $n \in \mathbb{N}$, $n \geq 2$, poniamo:

$$p_2(n) := \#\{(x, y) \in \mathbb{Z} \times \mathbb{Z} : \ x^2 + y^2 = n \ , \ \ \mathrm{MCD}(x, y) = 1 \, , \}$$

$${}_2p(n) := \#\{(x, y) \in \mathbb{Z} \times \mathbb{Z} : \ x^2 + y^2 = n \ , \ \ \mathrm{MCD}(x, y) = 1, \ x \geq 0, \ y \gneqq 0\} \, .$$

Mostrare che $p_2(2) = 4$ e, per $n \geq 3$, $p_2(n)$ $(= 4 \cdot {}_2p(n))$ è uguale a quattro volte il numero delle soluzioni della congruenza $X^2 \equiv -1 \pmod{n}$ e cioè:

$$p_2(n) = \begin{cases} 0 & \text{se } 4 \mid n \text{ oppure se qualche primo } p \equiv 3 \pmod 4 \text{ divide } n, \\ 4 \cdot 2^s & \text{se } 4 \nmid n \text{ e nessun primo } p \equiv 3 \pmod 4 \text{ divide } n\ , \end{cases}$$

essendo s il numero di divisori primi dispari distinti di n (cfr. Teorema 3.7).

(c) Mostrare che ${}_2p$ è una funzione (aritmetica) moltiplicativa, mentre p_2 è una funzione aritmetica che non è in generale moltiplicativa,

(d) Per ogni primo p, mostrare che $p_2(p) = r_2(p)$, e cioè:

$$p_2(p) = \begin{cases} 4 & \text{se } p = 2\,, \\ 8 & \text{se } p \equiv 1 \pmod 4\,, \\ 0 & \text{se } p \equiv 3 \pmod 4\,. \end{cases}$$

(e) Mostrare che ${}_2r(n) = \sum_{d^2 \mid n} {}_2p(n/d^2)$.

[*Suggerimento.* (a) Si noti, innanzitutto, che adattando opportunamente la dimostrazione, si può generalizzare il Lemma 3.5 nella forma seguente: *se* $a, n \geq 1$ *e se* $\mathrm{MCD}(a, n) = 1$, *allora la congruenza lineare in due intederminate*

$$aX \equiv Y \pmod n$$

ammette sempre una soluzione (x_0, y_0) *con* $0 < |x_0| < \sqrt{n}$ *e* $0 < |y_0| < \sqrt{n}$.

Si supponga che a sia una soluzione di $X^2 \equiv -1 \pmod n$ con $1 \leq a \leq n-1$. Si prenda (x_0, y_0) come nel precedente enunciato e si ponga $x := |x_0|$, $y := |y_0|$. Essendo $y_0^2 \equiv a^2 x_0^2 \pmod n$, quindi si ha che $x^2 + y^2 = x_0^2 + y_0^2 \equiv 0 \pmod n$ con $x > 0$ e $y > 0$. Essendo $0 < x, y < \sqrt{n}$, allora necessariamente $x^2 + y^2 = n$. Se x_0 ed y_0 hanno segni concordi, allora risulta anche $ax \equiv y \pmod n$. Se x_0 ed y_0 hanno segni discordi (ad esempio, per fissare le idee, $x_0 < 0$ ed $y_0 > 0$), allora poiché $a^2 \equiv -1 \pmod n$ risulta $-ay_0 \equiv x_0 \pmod n$. In tal caso, si prende $x' := |y_0|$ e $y' := |x_0| = x_0$ ed allora si avrà $x'^2 + y'^2 = x_0^2 + y_0^2 \equiv 0 \pmod n$, $x' > 0$, $y' > 0$, ed anche $ax' \equiv y' \pmod n$. Essendo $0 < x', y' < \sqrt{n}$, allora necessariamente $x'^2 + y'^2 = n$.

Vogliamo ora dimostrare che $\mathrm{MCD}(x, y) = 1$ (e quindi ovviamente anche $\mathrm{MCD}(x', y') = 1$).

Siano $h, k \in \mathbb{Z}$ tali che:

$$a^2 = -1 + kn \quad \text{e} \quad y = ax + hn$$

allora si vede facilmente che:

$$n = x^2 + y^2 = x^2 + (ax + hn)^2 = x^2(1 + a^2) + axhn + hn(ax + hn) = n(x(kx + ha) + yh)$$

e, quindi, che $x(kx + ha) + yh = 1$. Questo fatto implica che $\mathrm{MCD}(x, y) = 1$.

Per quanto riguarda l'unicità, siano (x_1, y_1), (x_2, y_2) due coppie di interi che verificano le condizioni (i) ed (ii), allora è subito visto (applicando la Proposizione 3.1) che

$$n^2 = (x_1x_2 + y_1y_2)^2 + (x_1y_2 - x_2y_1)^2 \ .$$

Pertanto, $0 < x_1x_2 + y_1y_2 \leq n$ ed inoltre:

$$x_1x_2 + y_1y_2 \equiv x_1x_2 + (ax_1)(ax_2) = x_1x_2 + a^2x_1x_2 \equiv 0 \pmod{n} \ .$$

Da ciò si ricava che $x_1x_2+y_1y_2 = n$ e, dunque, $x_1y_2-y_1x_2 = 0$. Poiché $\mathrm{MCD}(x_1, y_1) = 1 = \mathrm{MCD}(x_2, y_2)$, allora si deduce che $x_1 = x_2$ e $y_1 = y_2$.

Viceversa, sia (x, y) una coppia di interi verificante le condizioni (i) e (ii). Dalla relazione $ax \equiv y \pmod{n}$ e dal fatto che $\mathrm{MCD}(x, n) = 1$, si ricava che a è univocamente determinato $(\mathrm{mod}\ n)$, essendo $a \equiv x^*y \pmod{n}$. Inoltre, essendo $0 \equiv x^2 + y^2 \equiv x^2(1+a^2) \pmod{n}$, si ricava immediatamente che $a^2 \equiv -1 \pmod{n}$.

(b) Utilizzando (a), ${}_2p(n)$ coincide con il numero delle soluzioni incongruenti di $X^2 \equiv -1 \pmod{n}$. Tale numero è stato calcolato nel Corollario I.6.37.

(c) segue facilmente da (b).

(d) è una verifica diretta (cfr. anche Esercizio 3.7 (e))

(e) segue per verifica diretta, utilizzando il fatto che ${}_2p(n)$ e ${}_2r(n)$ sono funzioni moltiplicative.]

3.9. Siano $a, b \in \mathbb{Z}$. Supponiamo che se p è un primo dispari e se $p \mid a$ allora $p \equiv 1 \pmod{4}$.

(a) Mostrare che l'equazione diofantea in due indeterminate:

$$Y^2 + 4a^2 = X^3 + (4b-1)^3$$

non ha soluzioni.

(b) Mostrare che l'equazione diofantea $Y^2 = X^3 + 11$ non ha soluzioni.

[*Suggerimento*. (a) Sia per assurdo (x, y) una soluzione della equazione diofantea, allora $y^2 \equiv x^3 - 1 \pmod{4}$. Dal momento che $y^2 \equiv 0, 1 \pmod{4}$, allora necessariamente $x \equiv 1 \pmod{4}$. Poiché:

$$x^3 + (4b-1)^3 = (x^2 - x(4b-1) + (4b-1)^2)(x + (4b-1))$$

e poiché $x^2 - x(4b-1) + (4b-1)^2 \equiv 3 \pmod{4}$, deve esistere un primo $q \equiv 3 \pmod{4}$ tale che $q \mid (y^2 + 4a^2)$. Per il Corollario 3.9 (b), q deve allora dividere sia y che $2a$ e, quindi, q deve dividere a.

(b) Si sommi 16 ad ambo i membri dell'equazione data e si applichi il caso (a).]

3.10. Siano $a, b \in \mathbb{Z}$. Supponiamo che se p è un primo dispari e se $p \mid (2a+1)$ allora $p \equiv 1 \pmod{4}$.

(a) Mostrare che l'equazione diofantea in due indeterminate:

$$Y^2 + (2a+1)^2 = X^3 + (4b+2)^3$$

non ha soluzioni.

(b) Mostrare che la seguente equazione diofantea:

$$Y^2 = X^3 - 17$$

non ha soluzioni.

[*Suggerimento*. (a) Come sopra, se (x, y) è una soluzione allora $x \equiv 1 \pmod 4$ e $x + (4b + 2) \equiv 3 \pmod 4$, da cui si trova un primo $q \equiv 3 \pmod 4$ tale che $q \mid (y^2 + (2a + 1)^2)$ e quindi $q \mid (2a + 1)$.

(b) Si sommi 25 ad ambo i membri dell'equazione.]

Osservazione (Esercizi 3.9 e 3.10). L'equazione diofantea in due indeterminate:

$$Y^2 = X^3 + k \quad \text{con } k \in \mathbb{Z} \tag{3.10.1}$$

è stata a lungo studiata da molti matematici, tra i quali L.J. Mordell attorno al 1913.

Si noti che l'equazione (3.10.1) determina un esempio importante di curva ellittica. Si noti, inoltre, che tale equazione, per $k = -2$, fu considerata da Bachet nel 1621. Fermat affermò che tale equazione diofantea ha soltanto la soluzione $(3, \pm 5)$, ma la sua dimostrazione non fu mai pubblicata (cfr. Esercizio 3.11).

Mordell ha dimostrato alcuni risultati più generali di quelli enunciati negli esercizi precedenti. Si ponga $k = \beta^3 - \alpha^2$, allora l'equazione diofantea (3.10.1) non ha soluzioni nei seguenti casi:

(1) β dispari, α pari, $3 \nmid \alpha$, $p \mid \text{MCD}(\alpha, \beta) \Rightarrow p \equiv 1 \pmod 4$, $k \not\equiv 7 \pmod 8$
(ad esempio: $k = 13\ (= 17^3 - 70^2)$, $11\ (= 3^3 - 4^2)$, $-3\ (= 1^3 - 2^2)$, $-5\ (= (-1)^3 - 2^2)$);

(2) $\beta \equiv 2 \pmod 4$, α dispari, $3 \nmid \alpha$, $p \mid \text{MCD}(\alpha, \beta) \Rightarrow p \equiv 1 \pmod 4$
(ad esempio: $k = 7\ (= 2^3 - 1^2)$; $-9\ (= (-2)^3 - (1)^2)$);

(3) $\beta = 2b$, $\alpha = 2a$, con a dispari, $3 \nmid a$, $b \equiv 3 \pmod 4$, $p \mid \text{MCD}(a, b) \Rightarrow p \equiv 1 \pmod 4$
(ad esempio: $k = 20\ (= 6^3 - 14^2)$, $-12\ (= (-2)^3 - 2^2)$).

Successivamente A. Thue nel 1917 e L.J. Mordell nel 1922 hanno dimostrato che, per ogni $k \neq 0$, (3.10.1) ha al più un numero finito di soluzioni negli interi. Tuttavia, il numero delle soluzioni di (3.10.1) può essere arbitrariamente grande. Non è nota nessuna condizione generale per la risolubilità di (3.10.1) negli interi.

Come nel caso intero, non è noto per quali interi k l'equazione di Mordell (3.10.1) abbia soluzioni razionali. Importanti risultati sono stati ottenuti su tale problematica da Mordell (1969), Cassels (1950), Birch e Swinnerton–Dyer (1963). Il caso razionale differisce sostanzialmente dal caso intero per quanto riguarda il numero delle soluzioni. Infatti, Fueter nel 1930 ha dimostrato che se $k \neq 1, -432,$

se k non possiede fattori con potenza sei, e se (3.10.1) ha una soluzione razionale (x, y), con $xy \neq 0$, allora l'equazione (3.10.1) ha infinite soluzioni razionali. (Se $k = 1$, le soluzioni di (3.10.1) sono soltanto le seguenti $(0, \pm 1)$, $(-1, 0)$, $(2, \pm 3)$; se $k = -432$, l'equazione di Mordell ha un'unica soluzione $(12, \pm 36)$.)

Per maggiori dettagli sull'equazione diofantea di Mordell rinviamo a [9, Chapter 26] e [12, Section 14.4].

3.11. (Fermat, 1658) Mostrare che l'equazione diofantea

$$Y^2 = X^3 - 2$$

ha solamente le soluzioni intere non banali $x = 3$, $y = \pm 5$.

[*Suggerimento*. Si pensi di poter operare nell'anello $\mathbb{Z}[i\sqrt{2}]$ (invece che in $\mathbb{Z}$) che è noto essere anch'esso un dominio euclideo (e quindi, in particolare, un dominio a fattorizzazione unica). Allora, se $x, y \in \mathbb{Z}$ è una soluzione dell'equazione diofantea data, si ha

$$x^3 = (y + i\sqrt{2})(y - i\sqrt{2}) \ .$$

Non è difficile verificare che $\mathrm{MCD}_{\mathbb{Z}[i\sqrt{2}]}(y + i\sqrt{2}, y - i\sqrt{2}) = 1$.

Infatti, se $\alpha := a + bi\sqrt{2} \in \mathbb{Z}[i\sqrt{2}]$ divide $\beta := y + i\sqrt{2}$ e $\overline{\beta} = y - i\sqrt{2}$, allora esso dovrebbe dividere anche la loro differenza (cioè, $2i\sqrt{2}$) e la loro somma (cioè, $2y$). Quindi $N(\alpha) \mid N(2i\sqrt{2})$ e $N(\alpha) \mid N(2y)$, cioè $N(\alpha) \mid 8$ e $N(\alpha) \mid 4y^2$, dunque $N(\alpha) \mid 4$. Da cui si ricava che $a = \pm 1$ e $b = 0$ oppure $a = 0$ e $b = \pm 1$. È subito visto che nessuna di tali soluzioni determina un divisore proprio di β e $\overline{\beta}$. Pertanto, gli elementi β e $\overline{\beta}$ sono relativamente primi in $\mathbb{Z}[i\sqrt{2}]$.

Essendo $x^3 = \beta\overline{\beta}$ allora necessariamente deve esistere $\gamma := c + id\sqrt{2} \in \mathbb{Z}[i\sqrt{2}]$ in modo tale che $\beta = \gamma^3$. Dunque, confrontando i coefficienti di $i\sqrt{2}$, deve essere

$$1 = d(3c^3 - 2d^2)$$

pertanto $d = 1$ e $c = \pm 1$. Da cui si ricava che:

$$y + i\sqrt{2} = (\pm 1 + i\sqrt{2})^3$$

ovvero che $y = \pm 5$ e dunque $x = 3$.]

4 Interi somma di più di due quadrati

Abbiamo già osservato, risolvendo l'equazione diofantea $X^2 + Y^2 = n$, che *non* ogni intero positivo si può scrivere come somma di due quadrati di interi (ad esempio: $3 = 1^2+1^2+1^2$, $6 = 2^2+1^2+1^2$, $15 = 3^2+2^2+1^2+1^2$). Si pone naturalmente il problema di trovare il numero minimo di quadrati di interi (relativi o, equivalentemente, non negativi) necessari in una somma per scrivere un qualunque intero positivo.

Per quanto riguarda il problema degli interi positivi, che si possono scrivere come somma di tre quadrati, ma non di due (ad esempio: $3 = 1^2+1^2+1^2$, $43 = 3^2+3^2+5^2$), possiamo senz'altro dire che esso è notevolmente più difficile di quello corrispondente per due o quattro quadrati. Una delle ragioni è che, a differenza di quanto accade nel caso della somma di due (o come vedremo anche nel caso di quattro quadrati), il prodotto di due interi che si possono scrivere come somma di tre quadrati non è, in generale, un intero somma di tre quadrati. (Ad esempio, $3 \cdot 21 = (1^2+1^2+1^2)(4^2+2^2+1^2) = 63 = 7^2+3^2+2^2+1^2$ oppure $3 \cdot 13 = (1^2+1^2+1^2)(3^2+2^2+0^2) = 39 = (6^2+1^2+1^2+1^2)$; mentre, al contrario, $3 \cdot 43 = 129 = 11^2+2^2+2^2$ è ancora ottenibile come somma di tre quadrati.)

A proposito di tale problema segnaliamo che, nel 1785, Legendre affermò che *ogni intero positivo od il suo doppio si può scrivere come somma di tre quadrati di interi*, risultato che egli dimostrò poi completamente nel 1798 come conseguenza del suo teorema generale che asserisce che *ogni intero positivo che non è del tipo* $4k$*, né del tipo* $8k+7$*, è somma di tre quadrati di interi*. Il risultato più generale concernente tale problema è il seguente teorema dimostrato da Gauss, il quale ha completato il lavoro iniziato da Legendre; i successivi contributi di Cauchy e Dirichlet riguardano semplificazioni della dimostrazione di Gauss.

Teorema 4.1. (A.M. Legendre, 1798; K.F. Gauss, 1801) *Un numero naturale* $n \geq 1$ *può essere scritto come somma di tre quadrati di interi se, e soltanto se,* $n \neq 4^e(8k+7)$*, con* e, k *interi,* $e, k \geq 0$.

Dimostrazione. Ci limitiamo a dimostrare la parte "soltanto se" di tale teorema, parte che non presenta particolari difficoltà.

Sia $n = x^2 + y^2 + z^2$ un numero intero, con $n \geq 1$. Per mostrare che $n \neq 4^e(8k+7)$, ragioniamo per induzione sull'intero $e \geq 0$. Dal momento che il quadrato di un intero è congruo a 0, 1 e 4 (mod 8), allora n è congruo a $0, 1, 2, 3, 4, 5,$ o 6 (mod 8), cioè $n \neq 8k+7$, per ogni intero $k \geq 0$. Supponiamo

che fissato $e \geq 1$, $n \neq 4^{e-1}(8k+7)$, per ogni intero $k \geq 0$. Vogliamo mostrare che $n \neq 4^e(8k+7)$, per ogni intero $k \geq 0$. Se per assurdo $n = 4^e(8k+7)$, per qualche $k \geq 0$, allora $4 \mid n$, da cui x, y, z devono essere tutti pari e quindi $n/4 = 4^{e-1}(8k+7) = (x/2)^2 + (y/2)^2 + (z/2)^2$, e ciò è assurdo per l'ipotesi induttiva.

□

Osservazione 4.2. **(a)** Per la dimostrazione della parte "se" cfr., ad esempio, E. Landau [7] oppure L.J. Mordell [9].

(b) Se n è un numero naturale, $n \geq 1$, si può dimostrare che:

(1) se $n \equiv 0 \pmod 8$, allora, per infiniti valori di n (ad esempio, $n := 24 \cdot k^2$, $k \geq 1$), n può essere scritto come somma di tre quadrati di interi positivi ($24 \cdot k^2 = (4k)^2 + (2k)^2 + (2k)^2$) e, per infiniti valori di n (ad esempio, $n := 2^k$, per $k \geq 1$; risultato dimostrato da A. Hurwitz nel 1907, cfr. Esercizio 4.3 (b)), n *non* può essere scritto come somma di tre quadrati di interi positivi (cfr. Sierpiński [13, p. 406]);

(2) se $n \equiv 1 \pmod 8$, allora B. Jones e G. Pall nel 1939 hanno dimostrato che, tranne 1 e 25, tutti i numeri naturali di questo tipo si possono scrivere come somma di tre quadrati interi.

(3) se $n \equiv 4 \pmod 8$, allora $n = 8k + 4$ è somma di tre quadrati di interi positivi se, e soltanto se, $2k + 1$ gode della stessa proprietà (cfr. anche il successivo Esercizio 4.3 (a)).

(4) Se $n \equiv 6 \pmod 8$, allora n è somma di tre quadrati di interi (semplice conseguenza del Teorema 4.1), ma non è somma di due quadrati di interi, perché $n = 8k + 6 = 2(4k + 3)$.

Dal Teorema 4.1 discende che $7, 15, 23, 28, \ldots$ non sono somma di tre quadrati. Nel 1621, Bachet, nelle sue note alla *Arithmetica* di Diofanto, congetturò che ogni intero positivo si può scrivere come somma di quattro quadrati di interi, affermando tra l'altro di aver verificato tale congettura per tutti gli interi positivi $n \leq 325$. Fermat, in un'altra celebre annotazione, affermò di aver provato la Congettura di Bachet, usando il metodo della discesa infinita (cfr. la dimostrazione del successivo Teorema 4.7). Più tardi, egli sfidò gli altri cultori di tali problematiche suoi contemporanei a dimostrare questo risultato, in modo da verificare di non aver sopravvalutato questa sua notevole scoperta. Euler tentò a lungo di dare una dimostrazione

di tale risultato, ma ottenne soltanto dei risultati parziali. Lagrange, nel 1770, fu il primo a dare una dimostrazione completa della Congettura di Bachet, riconoscendo ad Euler un notevole contributo di idee che egli utilizzò nella sua dimostrazione.

Per dimostrare il Teorema di Lagrange dei quattro quadrati, procediamo, inizialmente, in una maniera analoga a quella seguita nel problema della decomposizione di un intero positivo come somma di due quadrati di interi.

Proposizione 4.3. (L. Euler, 1748) *Siano $n, m \in \mathbb{N}^+$. Se n e m possono essere scritti come somma di quattro quadrati di interi, allora anche nm può essere scritto come somma di quattro quadrati di interi.*

Dimostrazione. L'enunciato è una immediata conseguenza della seguente identità

$$\begin{aligned}(a^2+b^2+c^2+d^2)(e^2+f^2+g^2+h^2) = \\ = (ae+bf+cg+dh)^2+(af-be+ch-dg)^2+ \\ +(ag-ce+df-bh)^2+(ah-de+bg-cf)^2\ .\end{aligned} \tag{4.3.1}$$

□

Osservazione 4.4. La verifica di una identità, in particolare della identità (4.3.1), consiste in un banale sviluppo dei calcoli indicati ad ambo i membri dell'uguaglianza. La scoperta della identità (4.3.1), dovuta ad Euler, è stata fondamentale importanza per la risoluzione del problema in esame. Si noti che questa identità non è, al contrario di quella simile relativa alla somma di due quadrati, una semplice conseguenza di proprietà inerenti i numeri complessi. Essa è, invece, conseguenza delle proprietà algebriche di una struttura algebrica più riposta: i quaternioni.

A proposito delle identità, che sono strumenti ricorrenti di indagine in teoria dei numeri, Littlewood ha acutamente osservato che "ogni identità è banale, se è scoperta da...qualcun altro".

Corollario 4.5. *Se ogni numero primo può essere scritto come somma di quattro quadrati di interi, allora ogni intero positivo può essere scritto come somma di quattro quadrati di interi.*

Dimostrazione. Immediata conseguenza della Proposizione 4.3.

□

Proposizione 4.6. *Per ogni primo p, esistono $x, y \in \mathbb{Z}$ in modo tale che*

$$x^2 + y^2 \equiv -1 \pmod{p} \ .$$

Dimostrazione. **Caso 1**: $p = 2$. Allora, basta porre $x = 1$, $y = 0$.

Caso 2: $p \equiv 1 \pmod 4$. Allora basta porre $y = 0$ e x uguale ad una soluzione della congruenza $X^2 \equiv -1 \pmod p$, esistente in quanto in tal caso $\left(\frac{-1}{p}\right) = 1$ (cfr. Proposizione I.6.6 (h)).

Caso 3: $p \equiv 3 \pmod 4$. Allora basta trovare un intero y tale che la congruenza

$$X^2 \equiv -(y^2 + 1) \pmod{p}$$

è risolubile, cioè, usando il simbolo di Legendre, un intero y tale che

$$\left(\frac{-(y^2+1)}{p}\right) = 1 \ .$$

Ma, poiché:

$$\left(\frac{-(y^2+1)}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{y^2+1}{p}\right) = -\left(\frac{y^2+1}{p}\right)$$

basta trovare un intero y tale che $\left(\frac{y^2+1}{p}\right) = -1$. Dal momento che tutti e soli gli interi a, tali che esiste $y \in \mathbb{Z}$ per cui $a \equiv y^2 \pmod p$ sono per definizione gli interi a tali che $\left(\frac{a}{p}\right) = 1$, allora basta determinare un intero $a \pmod p$ in modo tale che $\left(\frac{a}{p}\right) = 1$ e $\left(\frac{a+1}{p}\right) = -1$. È evidente che esistono interi che verificano tale proprietà (altrimenti, poiché $\left(\frac{1}{p}\right) = 1$, si avrebbe $\left(\frac{1+1}{p}\right) = \left(\frac{2}{p}\right) = \left(\frac{3}{p}\right) = \ldots = \left(\frac{p-1}{p}\right) = 1$) (cfr. anche l'Esercizio I.6.19).

□

Teorema 4.7. (J. L. Lagrange, 1770) *Sia n un intero positivo. Allora, n può essere scritto come somma di quattro quadrati di interi.*

Dimostrazione. Per il Corollario 4.5, ci si può limitare al caso in cui $n = p$, con p primo. Se $p = 2$, allora $2 = 1^2 + 1^2 + 0^2 + 0^2$, dunque si può supporre p primo dispari. Consideriamo la congruenza in due indeterminate:

$$X^2 + Y^2 \equiv -1 \pmod{p} \ .$$

Per la Proposizione 4.6, tale congruenza è sempre risolubile, quindi possiamo trovare $t, x, y \in \mathbb{Z}$, con $|x|, |y| \leq p/2$, in modo tale che

$$x^2 + y^2 + 1 = tp \ .$$

Abbiamo, dunque, mostrato che un multiplo di p è somma di quattro quadrati, cioè $tp = x^2 + y^2 + z^2 + w^2$, con $x, y \in \mathbb{Z}$, $z = 1$, $w = 0$. Vogliamo, ora, dimostrare che tale proprietà vale per $t = 1$. Sappiamo che:

$$0 \lneqq t = \frac{x^2 + y^2 + 1}{p} \leq \frac{(p/2)^2 + (p/2)^2 + 1}{p} = p/2 + 1/p \lneqq p \ .$$

Sia k il più piccolo intero positivo tale che $kp = x_1^2 + y_1^2 + z_1^2 + w_1^2$, con $x_1, y_1, z_1, w_1 \in \mathbb{Z}$, e supponiamo, per assurdo, che $1 \lneqq k$ ($\lneqq p$). Siano x_2, y_2, z_2, w_2 elementi di $S := \{0, +1, -1, +2, -2, \ldots\}$, sistema completo di residui (mod k) *minimo in valore assoluto*, tali che $x_1 \equiv x_2 \pmod{k}$, $y_1 \equiv y_2 \pmod{k}$, $z_1 \equiv z_2 \pmod{k}$, $w_1 \equiv w_2 \pmod{k}$. Si ha, ovviamente, che $|x_2| \leq k/2$, $|y_2| \leq k/2$, $|z_2| \leq k/2$, $|w_2| \leq k/2$. Per le scelte effettuate è chiaro, che non può essere:

$$x_2 \equiv y_2 \equiv z_2 \equiv w_2 \equiv 0 \pmod{k} \ ,$$

perché, altrimenti, si avrebbe $kp \equiv 0 + 0 + 0 + 0 = 0 \pmod{k^2}$, da cui si avrebbe che $k \mid p$, donde un assurdo. Inoltre, non può essere

$$|x_2| = |y_2| = |z_2| = |w_2| = k/2 \ ,$$

perché altrimenti, si avrebbe $kp \equiv 4(k/2)^2 = k^2 \equiv 0 \pmod{k^2}$, da cui di nuovo un assurdo. Ora, risulta:

$$0 \equiv x_1^2 + y_1^2 + z_1^2 + w_1^2 \equiv x_2^2 + y_2^2 + z_2^2 + w_2^2 \pmod{k} \ ,$$

dunque, esisterà un intero t' in modo tale che:

$$x_2^2 + y_2^2 + z_2^2 + w_2^2 = t'k \quad \text{cioè} \quad t' = \frac{x_2^2 + y_2^2 + z_2^2 + w_2^2}{k} \ .$$

Per quanto sopra osservato, si ha:

$$1 \leq t' \lneqq \frac{(k/2)^2 + (k/2)^2 + (k/2)^2 + (k/2)^2}{k} = k \ .$$

D'altra parte, poiché kp e $t'k$ sono somma di quattro quadrati di interi, allora, per la Proposizione 4.3, anche il loro prodotto $k^2t'p$ è somma di quattro quadrati di interi, cioè per (4.3.1) risulta:

$$k^2t'p = a^2 + b^2 + c^2 + d^2 \ ,$$

con

$$\begin{aligned}
a &= x_1x_2 + y_1y_2 + z_1z_2 + w_1w_2 \equiv x_1^2 + y_1^2 + z_1^2 + w_1^2 \equiv 0 \pmod{k},\\
b &= x_1y_2 - y_1x_2 + z_1w_2 - w_1z_2 \equiv x_1y_1 - y_1x_1 + z_1w_1 - w_1z_1 = 0 \pmod{k},\\
c &= x_1z_2 - z_1x_2 + w_1y_2 - y_1w_2 \equiv x_1z_1 - z_1x_1 + w_1y_1 - y_1w_1 = 0 \pmod{k},\\
d &= x_1w_2 - w_1x_2 + y_1z_2 - z_1y_2 \equiv x_1w_1 - w_1x_1 + y_1z_1 - z_1y_1 = 0 \pmod{k}.
\end{aligned}$$

Quindi, $t'p = (a/k)^2 + (b/k)^2 + (c/k)^2 + (d/k)^2$, dove a/k, b/k, c/k, $d/k \in \mathbb{Z}$ e $1 \leq t' \lneqq k$, donde una contraddizione per la minimalità della scelta di k.

□

Osservazione 4.8. (Problema di Waring) Nello stesso periodo in cui Lagrange dimostrava il teorema dei quattro quadrati, il matematico inglese E. Waring (1734–1798), nel suo libro *Meditationes Algebraicae* (1770) congetturava il seguente teorema:

Per ogni esponente intero positivo s, esiste un intero positivo k tale che ogni intero positivo n è somma di k potenze s–esime di interi non negativi.

Denotiamo con $g(s)$ *il più piccolo intero positivo k (qualora esistente) tale che ogni numero naturale è la somma di k potenze s–esime di interi non negativi.* Allora, con altre parole, il teorema congetturato da Waring assicura l'esistenza di un $g(s)$, per ogni $s \geq 1$.

La congettura di Waring fu provata da D. Hilbert nel 1909. Notiamo che:

(a) per $s = 1$, il Teorema di Waring è banalmente vero;

(b) per $s = 2$, il Teorema di Lagrange e i risultati già esposti sugli interi somma di quadrati assicurano che $g(2) = 4$;

(c) per $s = 3$, Waring congetturò che $g(3) = 9$, risultato che è stato provato da A. Wieferich nel 1909;

(d) per $s = 4$, Waring congetturò che $g(4) = 19$. Tale congettura è stata dimostrata da R. Balasubramanian, F. Dress e J.–H. Deshonilles nel 1985;

(e) per $s = 5$, J.R. Chen nel 1964 ha dimostrato che $g(5) = 37$.

In generale, da alcuni lavori di L.E. Dickson (1913), S.S. Pillai (1936), K. Mahler (1957) e R.M. Stemmler (1964) si ricava che:

$$g(s) = 2^s + [(3/2)^2] - 2$$

per $5 \leq s \leq 200.000$ e definitivamente per s.

Limitiamoci qui a dimostrare la seguente disuguaglianza:

Proposizione 4.9. *Per ogni intero $s \geq 2$, risulta*

$$g(s) \geq 2^s + [(3/2)^s] - 2 \ .$$

Dimostrazione. Sia $n := 2^s[(3/2)^s] - 1$. Dal momento che, per ogni numero reale positivo α, si ha ovviamente che $[\alpha] \leq \alpha$, allora:

$$n \leq 2^s(3^s/2^s) - 1 \lneqq 3^s \ .$$

Inoltre, per definizione di $g(s)$, si deve avere che

$$n = x_1^s + \cdots + x_{g(s)}^s \quad \text{con } x_i \in \mathbb{Z} \ , \ x_i \geq 0 \ .$$

Poiché $n < 3^s$, necessariamente si deve avere $x_i \lneqq 3$, per ogni indice i, $1 \leq i \leq g(s)$. Sia $\gamma_2 := \#\{x_i : 1 \leq i \leq g(s), x_i = 2\}$; $\gamma_1 := \#\{x_i : 1 \leq i \leq g(s), x_i = 1\}$; $\gamma_0 := \#\{x_i : 1 \leq i \leq g(s), x_i = 0\}$. Dunque $\gamma_0 + \gamma_1 + \gamma_2 = g(s) \geq \gamma_1 + \gamma_2$ ed, inoltre, $n = 2^s\gamma_2 + \gamma_1 \geq 2^s\gamma_2$. Quindi:

$$n + 1 = 2^s[(3/2)^s] \geq 2^s\gamma_2 + 1$$

da cui $[(3/2)^s] - 1 \geq \gamma_2$. Dunque, si ha

$$\begin{aligned} g(s) \geq \gamma_1 + \gamma_2 &= (n - 2^s\gamma_2) + \gamma_2 = \\ &= n - (2^s - 1)\gamma_2 \geq n - (2^s - 1)([(3/2)^2] - 1) = \\ &= 2^s + [(3/2)^s] - 2 \ . \end{aligned}$$

□

Osservazione 4.10. Per ogni intero positivo s, è stato definito un altro intero, denotato con $G(s)$, che per molti aspetti è anche più interessante di $g(s)$. L'intero $G(s)$ è, per definizione, *il più piccolo intero positivo k tale che ogni numero naturale sufficientemente grande (cioè, ogni numero naturale con al più un numero finito di eccezioni) è la somma di k potenze s–esime di interi non negativi*. Evidentemente risulta:

(a) $G(2) = g(2) = 4$;

(b) $G(s) \leq g(s)$, per $s \geq 3$.

Inoltre, non è troppo difficile dimostrare che $G(s) \geq s + 1$, per ogni $s \geq 2$. In una serie di lavori, pubblicati tra il 1919 e il 1928 (apparsi sotto il titolo generale di "Some problems of *partitio numerorum*"), Hardy e Littlewood, usando metodi analitici, hanno determinato delle limitazioni superiori per $G(s)$. La più semplice, ma anche la più "grossolana", è la seguente:

$$G(s) \leq s2^{s-1} + 1$$

successivamente migliorata, tra l'altro, con la seguente:

$$G(s) \leq (s-2)2^{s-1} + 5 \ .$$

Risultati ancora più soddisfacenti sono stati ottenuti da I. Vinogradov nella seconda metà degli anni venti e da H. Heilbroun nel 1936, che ha migliorato, semplificando anche le dimostrazioni, quelli di Vinogradov.

Sorprendentemente, non è ancora noto il valore di $G(3)$, anche se nel 1942 Y.V. Linnik ha mostrato che $G(3) \leq 7$ e, per quanto visto sopra, $4 \leq G(3)$. D'altro canto è noto che $G(4) = 16$, $G(5) \leq 21$ e $G(6) \leq 31$.

Per ulteriori informazioni ed una dettagliata bibliografia sul problema di Waring, rinviamo ai volumi di Hardy–Wright [6, p. 335–339] e Sierpiński [13, p. 427–430].

4 Esercizi e complementi

4.1. Esprimere 247 e 308 come somma di tre quadrati.

[*Soluzione*: $247 = 13 \cdot 19$ con:

$$13 = 3^2 + 2^2 + 0^2 + 0^2 \ ,$$
$$19 = 4^2 + 1^2 + 1^2 + 1^2 \ ,$$

dunque, utilizzando (3.3.1), abbiamo:

$$247 = (12+2)^2 + (3-8)^2 + (3-2)^2 + (3+2)^2 = 14^2 + 5^2 + 1^2 + 5^2 \ .$$

Si noti che tale scrittura non è unica, utilizzando ad esempio il fatto che $13 = 2^2 + 2^2 + 2^2 + 1^2$.

Dopo aver osservato che $308 = 2^2 \cdot 7 \cdot 11$ e che $7 = 2^2 + 1^2 + 1^2 + 1^2$ e $11 = 3^2 + 1^2 + 1^2 + 0^2$, si può dedurre che $308 = 8^2 + 8^2 + 12^2 + 6^2$.]

4.2. Dimostrare il Teorema di Lagrange (Teorema 4.7) supponendo di aver già dimostrato il Teorema di Legendre–Gauss (Teorema 4.1) (ovviamente, questo non sarebbe stato possibile a Lagrange perché il Teorema 4.1 è stato dimostrato, da un punto di vista temporale, dopo il Teorema 4.7).

[*Suggerimento*. Basta dimostrare che ogni primo dispari è somma di quattro quadrati. Se $p \equiv 1 \pmod 4$ allora p si scrive come somma di due quadrati (e, quindi, banalmente come somma di quattro quadrati). Se $p \equiv 3 \pmod 4$ allora $p - 1 = 4k + 2$ per qualche $k \geq 0$. Dunque $p - 1 = 8h + 6$ oppure $p - 1 = 8h + 2$ per qualche $h \geq 0$, rispettivamente se k è dispari oppure se k è pari. In entrambi i casi, per il Teorema di Legendre–Gauss, $p - 1 = a^2 + b^2 + c^2$ e quindi $p = a^2 + b^2 + c^2 + 1^2$.]

4.3. Mostrare che:

(a) Un numero naturale $n = 4k$, $k \geq 1$, è somma di tre quadrati di interi positivi se, e soltanto se, k gode della medesima proprietà.

(b) **(A. Hurwitz, 1907)** Se $n = 2^k$, $k \geq 1$, allora n *non* si può scrivere come somma di tre quadrati di interi positivi.

[*Suggerimento*. (a) è una semplice conseguenza del Teorema 4.1. Oppure, si noti che:

$$4k = a^2 + b^2 + c^2 \Leftrightarrow k = a_1^2 + b_1^2 + c_1^2 \quad \text{con } a = 2a_1,\ b = 2b_1,\ c = 2c_1 \ .$$

dove $a_1,\ b_1,\ c_1 \in \mathbb{Z}$ se e soltanto se $a,\ b,\ c \in \mathbb{Z}$, perché il quadrato di un qualunque intero è congruo a $0,\ 1 \pmod 4$.

(b) Per $k = 1, 2$ è ovvio. Se $k \geq 3$, l'enunciato discende da (a) per induzione su k perché se $2^k = 4 \cdot 2^{k-2} = a^2 + b^2 + c^2$ allora anche 2^{k-2} sarebbe somma di tre quadrati di interi.]

4.4. Usando il Teorema di Legendre–Gauss (Teorema 4.1), provare che un numero naturale è somma di tre quadrati di numeri razionali se, e soltanto se, è somma di tre quadrati di interi.

[*Suggerimento*. Sia $n = \frac{a^2}{d^2} + \frac{b^2}{d^2} + \frac{c^2}{d^2}$ con $a, b, c, d \in \mathbb{Z}$ e $d \neq 0$. Dunque $nd^2 = a^2 + b^2 + c^2$. Sia, per assurdo, $n = 4^e(8k+7)$ con $e, k \geq 0$. Possiamo sempre porre:

$$d = 2^f(2h+1) \ , \quad \text{per qualche } f, h \geq 0$$

e dunque otteniamo:

$$nd^2 = 4^{e+f}(8t+7) \quad \text{dove } e+f, t \geq 0 \ .$$

Ciò contraddice il Teorema di Gauss perché nd^2 è somma di tre quadrati di interi.]

4.5. Mostrare che ogni intero non negativo n si può esprimere nella forma $a^2 \pm b^2 \pm c^2$ (infatti, nella forma $a^2 + b^2 \pm c^2$), per una oppurtuna scelta di $a, b, c \in \mathbb{Z}$.

[*Suggerimento*. Dato $n > 0$ è sempre possibile trovare $a \in \mathbb{Z}$ in modo tale che $n - a^2$ è un intero dispari positivo.
Infatti se a è il più grande intero non negativo tale che $n > a^2$ e se $n - a^2$ è pari, basta prendere $a' := a - 1$ ed allora $n - a'^2$ è dispari. Dunque, $n - a^2 = 2m + 1$ per qualche $a, m \geq 0$. Dalla identità

$$2m + 1 = (m+1)^2 - m^2$$

ricaviamo che $n = a^2 + (m+1)^2 - m^2$. Ad esempio, se $n = 11$, $11 - 3^2 = 2$ e $11 - 2^2 = 7$, quindi $7 = 2 \cdot 3 + 1$ e $11 = 2^2 + 4^2 - 3^2$.]

4.6. (L. Euler, 1749) Mostrare che se n si scrive come somma di quattro quadrati di interi dispari, allora n si può scrivere anche come somma di quattro quadrati di interi pari.

[*Suggerimento*. Sia $n = (2a+1)^2 + (2b+1)^2 + (2c+1)^2 + (2d+1)^2$. È subito visto che $(2a+1)^2 + (2b+1)^2 = 2[(a+b+1)^2 + (a-b)^2]$. Dunque

$$\begin{aligned} n &= 2\left[(a+b+1)^2 + (a-b)^2 + (c+d+1)^2 + (c-d)^2\right] = \\ &= 2\left[(2\alpha+1)^2 + 4\beta^2 + (2\gamma+1)^2 + 4\delta^2\right] \end{aligned}$$

dove $a+b+1 = 2\alpha+1$, $c+d+1 = 2\gamma+1$ sono dispari e $a-b = 2\beta$, $c-d = 2\delta$ sono pari. Quindi

$$n = 4[(\alpha+\gamma+1)^2 + (\alpha-\gamma)^2 + \beta^2 + \delta^2] \ . \quad]$$

5 La (così detta) equazione di Pell: $X^2 - dY^2 = 1$

Attorno al 1657, Fermat pose il seguente problema:

Dato comunque un intero positivo che non è un quadrato, mostrare che esistono infiniti numeri interi tali che se il quadrato di ognuno di essi è moltiplicato per il numero assegnato e se 1 *viene aggiunto al risultato, allora ciò che si ottiene è ancora il quadrato di un numero intero.*

In simboli, ciò si può tradurre semplicemente nella seguente maniera: *dato un intero positivo d, che non sia un quadrato, mostrare che esistono infinite soluzioni intere per l'equazione diofantea*:

$$X^2 - dY^2 = 1 \ . \tag{5.1}$$

Osservazione 5.1. (a) Come osservato da Brouncker e Wallis attorno al 1660, si vede agevolmente che l'equazione (5.1) possiede infinite soluzioni razionali. Infatti, ponendo $X = 1 + (m/n)Y$, con $m, n \in \mathbb{Z}$, $n \geq 0$, si ha infatti la seguente equazione quadratica in una indeterminata:

$$1 + \left(\frac{m}{n}\right)^2 Y^2 + 2\frac{m}{n}Y - dY^2 = 1 \ .$$

Pertanto una soluzione di tale equazione è data da $y = 0$ (e quindi $x = 1$); l'altra soluzione (al variare di m ed n) è data da:

$$y = 2mn/(n^2 d - m^2)$$

e quindi: $x = (n^2 d + m^2)/(n^2 d - m^2)$.

(b) A causa dell'estrema parsimonia di informazioni sulla propria attività matematica, non sembrano essere chiaramente note le motivazioni che avrebbero portato Fermat allo studio di un tale problema. Probabilmente, Fermat fu stimolato dalla lettura della *Arithmetica Infinitorum* di J. Wallis suo contemporaneo. È evidente però che tale questione non sia stata posta "a caso". Ad esempio, l'equazione (5.1) è di importanza fondamentale per il problema generale della risoluzione di una qualunque equazione diofantea quadratica in due variabili. Infatti, si può dimostrare che tutte le equazioni diofantee di secondo grado in due variabili possono essere "ricondotte" allo studio di una equazione del tipo (5.1), (cfr. anche l'Esercizio 5.8).

Si noti, inoltre, che riferimenti a casi specifici dell'equazione di Pell si incontrano sovente nella matematica classica: ad esempio, il cosiddetto "problema del bestiame" di Archimede (III sec. a.C.). Questo è un problema con

otto incognite (relative a tipi differenti di capi di bestiame) che soddisfano sette relazioni lineari e due condizioni che assicurano che alcuni numeri sono quadrati perfetti. Tramite tecniche di eliminazione di indeterminate, questo problema veniva ricondotto alla risoluzione di un'equazione diofantea del tipo $X^2 - 4729494 \cdot Y^2 = 1$.

(c) J. Wallis, come osservato in (a), fu tra i primi a dare un metodo per determinare delle soluzioni al problema di Fermat sopra enunciato, per valori assegnati di d, senza però dimostrare, in generale, che tale metodo potesse permettere di determinare tutte le soluzioni. Questa "parte rimanente" del problema è certamente di particolare difficoltà, in quanto lo stesso Euler, pur provandoci, non pervenne ad una conclusione definitiva.

Lagrange nel 1766 dette la prima dimostrazione completa del Teorema di Fermat enunciato all'inizio del presente paragrafo. Questo problema è, oggi, comunemente noto come "problema della risoluzione dell'equazione di Pell" (dal nome di un matematico inglese contemporaneo del Wallis). Ciò è dovuto molto probabilmente al fatto che Euler, nel 1730, (quando cioè aveva appena 23 anni) aveva avuto l'errata impressione, leggendo l'opera del Wallis, che questi attribuisse a Pell l'idea del metodo di risoluzione dell'equazione diofantea $X^2 - dY^2 = 1$. Il nome di "equazione di Pell" è poi restato ad indicare l'equazione (5.1), benché Pell non si sia mai occupato effettivamente di tale equazione. (Per maggiori informazioni storiche e bibliografiche sull'argomento, cfr. I.E. Dickson [4, vol. II, Ch. XII] ed anche il volume di E. Whitford [17] dedicato a tale argomento.)

Prima di passare a descrivere le tecniche che ci porteranno alla risoluzione dell'equazione (5.1), vogliamo giustificare la presenza dell'ipotesi fatta su d, e cioè che d non sia un quadrato (di un intero). Infatti, se $d = a^2$, con $a \in \mathbb{Z}$, $a \neq 0$, allora:

$$1 = X^2 - dY^2 = (X - aY)(X + aY)$$

ovvero: $X - aY = \pm 1$ e $X + aY = \pm 1$; ciò equivale a dire che l'equazione diofantea (5.1) ammette solamente le soluzioni banali $x = \pm 1$ e $y = 0$. Più generalmente, possiamo supporre che d sia un intero privo di fattori quadratici. Infatti, se fosse $d = a^2 d'$ con $a, d' \in \mathbb{Z}$ e d' privo di fattori quadratici, allora si avrebbe che (x_0, y_0) è una soluzione di $X^2 - dY^2 = 1$ se, e soltanto se, (x_0, ay_0) è una soluzione di $X^2 - d'Y^2 = 1$.

Infine, ovviamente, si suppone $d > 0$, poiché, se fosse $d < 0$, allora si avrebbe $x^2 - d^2y^2 \geq 0$ presi comunque $x, y \in \mathbb{Z}$, e dunque $x^2 - dy^2 = 1$, se, e soltanto se, $x = \pm 1$ e $y = 0$ ed anche, nel caso in cui $d = -1$, $x = 0$

e $y = \pm 1$. Quindi, per $d < 0$, l'equazione di Pell avrebbe soltanto soluzioni banali.

Procediamo, ora, alla dimostrazione dell'esistenza di infinite soluzioni dell'equazione diofantea di Pell $X^2 - dY^2 = 1$, quando d è *un intero positivo che non possiede fattori quadratici (interi)*.

Lemma 5.2. *Presi, comunque,* $x_1, y_1, x_2, y_2, d \in \mathbb{Z}$, *vale la seguente identità:*

$$(x_1^2 - dy_1^2)(x_2^2 - dy_2^2) = (x_1x_2 \mp dy_1y_2)^2 - d(x_1y_2 \mp y_1x_2)^2 \ ,$$

(dove la scelta dei segni a secondo membro deve essere concorde).

Dimostrazione. La verifica è diretta.

□

Corollario 5.3. *Se* (x_1, y_1), (x_2, y_2) *sono due soluzioni dell'equazione di Pell, allora anche le seguenti coppie:*

$$(x_3 := x_1x_2 - dy_1y_2,\ y_3 := x_1y_2 - y_1x_2)\ ,\ (x_4 := x_1x_2 + dy_1y_2,\ y_4 := x_1y_2 + y_1x_2)$$

sono soluzioni dell'equazione di Pell.

□

Lemma 5.4. *Sia* $\alpha \in \mathbb{R}$ *e sia* N *un intero positivo. Allora, esistono due interi* p *e* q *tali che* $1 \leq q \leq N$ *e*

$$|\alpha - p/q| \leq 1/qN \ .$$

Dimostrazione. La prova di questo lemma, che ci fornisce una "approssimazione razionale" conveniente per un qualunque numero reale (non razionale), è basata sul cosiddetto Principio di Dirichlet (cfr. il Paragrafo 3 di questo Capitolo). Sia $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. Si considerino i seguenti $N+1$ numeri reali

$$k\alpha - [k\alpha]\ , \quad 0 \leq k \leq N\ ,$$

compresi tra 0 e 1. Si divida l'intervallo reale $[0, 1]$ in N sub–intervallini uguali di ampiezza esattamente $\frac{1}{N}$ e cioè:

$$\left[\frac{h}{N}, \frac{h+1}{N}\right] , \quad 0 \leq h \leq N-1 .$$

Allora, per il Principio di Dirichlet, almeno due tra i numeri reali $k\alpha - [k\alpha]$ appartengono allo stesso intervallino, cioè esistono $n, m \in \mathbb{Z}$ con $0 \leq n < m \leq N$ tali che, se $q := m - n$ e $p := [m\alpha] - [n\alpha]$,

$$|q\alpha - p| \leq \frac{1}{N} .$$

□

Corollario 5.5. *Sia $\alpha \in \mathbb{R}$. Esistono infinite coppie di interi (p, q) con $q \geq 1$ tali che:*

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{q^2} .$$

Dimostrazione. Possiamo supporre che α sia irrazionale (cioè, $\alpha \in \mathbb{R} \setminus \mathbb{Q}$), altrimenti l'affermazione è banalmente soddisfatta. Ponendo $N := N_1 := 1$, per il Lemma 5.4 possiamo affermare che esistono $p_1, q_1 \in \mathbb{Z}$ (con $q_1 = 1$) tali che:

$$\left|\alpha - \frac{p_1}{q_1}\right| \leq \frac{1}{q_1 N_1} = \frac{1}{q_1^2} .$$

Poiché α è irrazionale, si ha che $|q_1\alpha - p_1| \neq 0$; pertanto si può trovare $1 \lneqq N_2$ tale che

$$\frac{1}{N_2} < |q_1\alpha - p_1| .$$

Per il Lemma 5.4, applicato al caso $N := N_2$, possiamo affermare che esistono q_2, $1 \leq q_2 \leq N_2$, e $p_2 \in \mathbb{Z}$ tali che:

$$\left|\alpha - \frac{p_2}{q_2}\right| \leq \frac{1}{q_2 N_2} \leq \frac{1}{q_2^2} .$$

Dunque, abbiamo:

$$|q_2\alpha - p_2| \leq 1/N_2 < |q_1\alpha - p_1| ,$$

quindi $(p_2, q_2) \neq (p_1, q_1)$. Il procedimento precedente si può iterare essendo $|q_2\alpha - p_2| \neq 0$.

□

Lemma 5.6. *Sia* $B := 2\sqrt{d} + 1$, *con* d *intero positivo privo di fattori quadratici. Allora, esistono infinite coppie di interi* (x, y) *tali che:*

$$|x^2 - dy^2| \leq B \ .$$

Dimostrazione. Per il Corollario 5.5, esistono infinite coppie di interi x, y con $y \geq 1$ tali che:

$$\left|\sqrt{d} - \frac{x}{y}\right| \leq \frac{1}{y^2} ,$$

da cui si ricava che:

$$|\sqrt{d}y - x| \leq \frac{1}{y} ,$$

ed anche che:

$$\left|\frac{x}{y}\right| \leq \sqrt{d} + \frac{1}{y^2} \leq \sqrt{d} + 1 \ .$$

Pertanto:

$$\begin{aligned} |x^2 - dy^2| &= |\sqrt{d}y + x| \cdot |\sqrt{d}y - x| \leq \left|\frac{\sqrt{d}y + x}{y}\right| = \left|\sqrt{d} + \frac{x}{y}\right| \leq \\ &\leq \sqrt{d} + \left|\frac{x}{y}\right| \leq \sqrt{d} + \sqrt{d} + 1 = B \ . \end{aligned}$$

□

Teorema 5.7. *Sia* d *un intero positivo privo di fattori quadratici. Allora, l'equazione di Pell:*

$$X^2 - dY^2 = 1 \tag{5.1}$$

ha infinite soluzioni distinte.

Dimostrazione. Per il Lemma 5.6, ci sono infinite coppie di interi (x, y) tali che $|x^2 - dy^2| \leq B = 2\sqrt{d} + 1$. Poiché $x^2 - dy^2$ è un intero e poiché ci sono solo un numero finito di interi k tali che $0 \leq |k| \leq B$, allora esiste un intero k_0, $-B \leq k_0 \leq B$, in corrispondenza del quale esistono infinite coppie di interi (x, y) tali che $x^2 - dy^2 = k_0$. Si noti che $k_0 \neq 0$ (altrimenti d sarebbe un quadrato perfetto).
Esaminiamo l'insieme delle soluzioni (x, y) dell'equazione diofantea $X^2 - dY^2 = k_0$ e consideriamole modulo $|k_0|$. Evidentemente ci sono solo k_0^2 coppie (a, b) tali che $x \equiv a \pmod{|k_0|}$, $y \equiv b \pmod{|k_0|}$ e $0 \leq a, b < |k_0|$, al variare di (x, y).

Poiché l'equazione $X^2 - dY^2 = k_0$ ammette infinite soluzioni, debbono esistere allora due interi a, b, con $0 \leq a,\, b < |k_0|$, in modo tale che le soluzioni (x, y) di $X^2 - dY^2 = k_0$ soddisfacenti simultaneamente anche il sistema di congruenze:

$$\begin{cases} X \equiv a \pmod{|k_0|} \\ Y \equiv b \pmod{|k_0|} \end{cases}$$

siano infinite. Presa comunque una coppia di tali soluzioni, siano esse (x_1, y_1) e (x_2, y_2) con $(x_1, y_1) \neq (x_2, y_2)$, allora:

$$\begin{aligned} &x_1^2 - dy_1^2 = k_0 = x_2^2 - dy_2^2\,, \\ &x_1 \equiv a \equiv x_2 \pmod{|k_0|}\,, \\ &y_1 \equiv b \equiv y_2 \pmod{|k_0|}\,, \end{aligned}$$

Usando, ora, l'identità del Lemma 5.2, si ha:

$$k_0^2 = (x_1x_2 - dy_1y_2)^2 - d(x_1y_2 - y_1x_2)^2\ .$$

Inoltre,

$$\begin{aligned} &x_1x_2 - dy_1y_2 \equiv x_1^2 - dy_1^2 = k_0 \equiv 0 \pmod{|k_0|}\,, \\ &x_1y_2 - y_1x_2 \equiv x_1y_1 - y_1x_1 = 0 \pmod{|k_0|}\,. \end{aligned}$$

Perciò:

$$1 = \left(\frac{x_1x_2 - dy_1y_2}{k_0}\right)^2 - d\left(\frac{x_1y_2 - y_1x_2}{k_0}\right)^2$$

e quindi la coppia:

$$\left(x_3 := \frac{x_1x_2 - dy_1y_2}{k_0}\ ,\ y_3 := \frac{x_1y_2 - y_1x_2}{k_0}\right)$$

è una soluzione dell'equazione di Pell. Ora, se fissiamo (x_1, y_1) (si noti che $k_0 \neq 0$, e che x_1 e y_1 non sono entrambi nulli) e facciamo variare (x_2, y_2), dal momento che x_2 o y_2 possono assumere infiniti valori distinti, abbiamo trovato infinite soluzioni distinte (descritte dalla coppia (x_3, y_3)) per l'equazione di Pell.

□

Lemma 5.8. *Siano $x_1, y_1, x_2, y_2, d \in \mathbb{Z}$. Allora, valgono le seguenti identità:*

$$\begin{aligned} &(x_1 + \sqrt{d}y_1)(x_2 + \sqrt{d}y_2) = (x_1x_2 + dy_1y_2) + \sqrt{d}(x_1y_2 + y_1x_2)\,, \\ &(x_1 + \sqrt{d}y_1)(x_2 - \sqrt{d}y_2) = (x_1x_2 + dy_1y_2) - \sqrt{d}(x_1y_2 + y_1x_2)\,. \end{aligned}$$

Dimostrazione. La verifica è banale.

□

Corollario 5.9. *Siano $x_1, y_1, d, n \in \mathbb{Z}$, $n \geq 1$. Allora:*

$$(x_1 + \sqrt{d}y_1)^n = x_n + \sqrt{d}y_n \,,$$
$$(x_1 - \sqrt{d}y_1)^n = x_n - \sqrt{d}y_n \,,$$

con $x_n := x_1 x_{n-1} + d y_1 y_{n-1}$ e $y_n := x_1 y_{n-1} + y_1 x_{n-1}$, per $n \geq 2$.

Dimostrazione. Basta ragionare per induzione su n, ed applicare il Lemma 5.8.

□

Lemma 5.10. *Sia $n \geq 1$ e sia (x_1, y_1) una soluzione dell'equazione di Pell. Allora, se definiamo (come nel Corollario 5.9.) x_n e y_n tramite la relazione*

$$(x_1 + \sqrt{d}y_1)^n = x_n + \sqrt{d}y_n$$

allora anche (x_n, y_n) è una soluzione dell'equazione di Pell.

Dimostrazione.

$$\begin{aligned} x_n^2 - d y_n^2 &= (x_n + \sqrt{d}y_n)(x_n - \sqrt{d}y_n) = (x_1 + \sqrt{d}y_1)^n (x_1 - \sqrt{d}y_1)^n = \\ &= ((x_1 + \sqrt{d}y_1)(x_1 - \sqrt{d}y_1))^n = (x_1^2 - d y_1^2)^n = 1 \,. \end{aligned}$$

□

Osservazione 5.11. Sia (x_1, y_1) una soluzione dell'equazione di Pell. Essendo

$$(x_1 + \sqrt{d}y_1)^n (x_1 - \sqrt{d}y_1)^n = (x_1^2 - d y_1^2)^n = 1 \,,$$

allora, per ogni $n \geq 1$, si ha che l'inverso di $(x_1 + \sqrt{d}y_1)^n$ è dato da:

$$(x_1 + \sqrt{d}y_1)^{-n} = (x_1 - \sqrt{d}y_1)^n = x_n - \sqrt{d}y_n \,.$$

Si noti, anche, che per per ogni $n \geq 1$,

$$x_n + \sqrt{d}y_n = (x_n - \sqrt{d}y_n)^{-1} \,.$$

Osserviamo che se (x, y) è una soluzione dell'equazione di Pell, allora anche $(\pm x, \pm y)$ è una soluzione, per ogni possibile scelta di segno; quindi, basta determinare tutte le soluzioni di (5.1) per le quali $x \geq 0$ ed $y \geq 0$. È evidente poi che le soluzioni per le quali x o y è zero sono solo $(\pm 1, 0)$; quindi, in definitiva, basta determinare tutte le soluzioni per le quali $x > 0$ e $y > 0$. Tali soluzioni sono chiamate *soluzioni positive dell'equazione di Pell.*

Proposizione 5.12. *Sia (x, y) una soluzione dell'equazione di Pell. Allora, (x, y) è una soluzione positiva se, e soltanto se, $x + \sqrt{d}y > 1$.*

Dimostrazione. È chiaro che, se $x \geq 1$ e $y \geq 1$, allora $x + \sqrt{d}y \geq 1 + \sqrt{d} \geq 2 > 1$.

Viceversa, supponiamo che $x + \sqrt{d}y > 1$; quindi $(x, y) \neq (\pm 1, 0)$. Esaminiamo le varie possibilità:

Caso 1: Se $x < 0$ e $y < 0$, allora anche $x + \sqrt{d}y < 0$, donde un assurdo.

Caso 2: Se $x > 0$ e $y < 0$, allora $x - \sqrt{d}y \geq 1 + \sqrt{d} \gneqq 1$, quindi $1 = x^2 - dy^2 = (x + \sqrt{d}y)(x - \sqrt{d}y)$ è assurdo.

Caso 3: Se $x < 0$ e $y > 0$, allora $-x + y\sqrt{d} > 1$, quindi $-1 = -x^2 + dy^2 = (-x + \sqrt{d}y)(x + \sqrt{d}y)$ è assurdo.

Caso 4: Il caso $x > 0$ e $y > 0$ è pertanto l'unico caso possibile affinché risulti $x + \sqrt{d}y > 1$.

□

Definizione 5.13. La soluzione positiva (x_1, y_1) della equazione (5.1) per la quale $x_1 + \sqrt{d}y_1$ è minimo è detta *soluzione fondamentale dell'equazione di Pell.*

Si noti che la definizione sopra data è ben posta. Sia infatti (x_0, y_0) una soluzione positiva dell'equazione di Pell, che sappiamo esistere (Teorema 5.7). Per la Proposizione 5.12, $M := x_0 + \sqrt{d}y_0 > 1$. Se (x, y) è una qualunque altra soluzione positiva tale che $x + \sqrt{d}y \leq M$, allora $x \leq M$ e $y \leq M$. Pertanto, ci sono soltanto un numero finito di scelte intere (positive) per x e y. Quindi, è possibile trovare soltanto un numero finito di numeri reali del tipo $1 < x + \sqrt{d}y \leq M$.

Teorema 5.14. *Sia $d > 0$, d privo di fattori quadratici. Sia (x_1, y_1) la soluzione fondamentale dell'equazione di Pell:*

$$X^2 - dY^2 = 1 \tag{5.1}$$

e sia (x_n, y_n), $n \geq 1$, *tale che* $(x_1 + \sqrt{d}y_1)^n = x_n + \sqrt{d}y_n$ *(Corollario 5.9). Allora, tutte e sole le soluzioni dell'equazione di Pell sono date da:*

$$(\pm 1, 0) \quad \text{e} \quad (\pm x_n, \pm y_n)\,, \qquad n \geq 1\,,$$

dove tutte le possibilità di scelta per i segni sono ammesse. Inoltre, tutte queste soluzioni sono distinte tra loro.

Dimostrazione. Già sappiamo che $(\pm x_n, \pm y_n)$ e $(\pm 1, 0)$ sono soluzioni dell'equazione di Pell. Inoltre, $x_1 > 0$ e $y_1 > 0$ implica che $x_n > 0$ e $y_n > 0$; pertanto ogni soluzione del tipo $(\pm x_n, \pm y_n)$ è distinta dalle soluzioni $(\pm 1, 0)$. Per mostrare che le soluzioni $(\pm x_n, \pm y_n)$ sono tutte distinte, basta provare che $(x_n, y_n) \neq (x_m, y_m)$ per $n \neq m$. Supponiamo, per assurdo, che $(x_n, y_n) = (x_m, y_m)$ con $n \lneqq m$; allora dalla uguaglianza:

$$(x_1 + \sqrt{d}y_1)^n = x_n + \sqrt{d}y_n = x_m + \sqrt{d}y_m = (x_1 + \sqrt{d}y_1)^m$$

segue che $(x_1 + \sqrt{d}y_1)^{m-n} = 1$ con $m - n > 0$. Ciò è assurdo, poiché, essendo (x_{m-n}, y_{m-n}) una soluzione positiva dell'equazione di Pell, per la Proposizione 5.12 si ha che $(x_1 + \sqrt{d}y_1)^{m-n} \gneqq 1$.

Mostriamo, ora, che ogni altra soluzione (u, v) dell'equazione di Pell deve coincidere con una delle soluzioni sopra descritte. Possiamo limitarci ovviamente a supporre che $u \gneqq 0$ e $v \gneqq 0$. Per come si è scelta la soluzione fondamentale si ha che:

$$x_1 + \sqrt{d}y_1 \leq u + \sqrt{d}v\ .$$

Affermiamo che esiste un intero $n > 0$ tale che

$$(x_1 + \sqrt{d}y_1)^n \leq u + \sqrt{d}v \lneqq (x_1 + \sqrt{d}y_1)^{n+1}\ .$$

Infatti, essendo $x_1 + \sqrt{d}y_1 > 1$, il numero reale $(x_1 + \sqrt{d}y_1)^k$ cresce al crescere di k; basta quindi considerare il più grande intero positivo n tale che $(x_1 + \sqrt{d}y_1)^n \leq u + \sqrt{d}v$. Moltiplicando la disuguaglianza precedente per $(x_1 - \sqrt{d}y_1)^n$ e tenendo conto che $x_1 - \sqrt{d}y_1 > 0$, in quanto $x_1 + \sqrt{d}y_1 > 1$ e $(x_1 + \sqrt{d}y_1)(x_1 - \sqrt{d}y_1) = 1$, si ottiene:

$$1 \leq (u + \sqrt{d}v)(x_1 - \sqrt{d}y_1)^n < x_1 + \sqrt{d}y_1\ .$$

Se poniamo $u_1 := ux_n - dvy_n$ e $v_1 := vx_n - y_n u$, allora

$$1 \leq u_1 + \sqrt{d}v_1 < x_1 + \sqrt{d}y_1\ .$$

D'altra parte, si verifica subito che

$$u_1^2 - dv_1^2 = (u^2 - dv^2)(x_n^2 - dy_n^2) = 1 \cdot 1 = 1 \ ,$$

cioè, che (u_1, v_1) è una soluzione dell'equazione di Pell. Quindi, essendo (x_1, y_1) la soluzione fondamentale, si ha che $u_1 + \sqrt{d}v_1 = 1$. Quindi,

$$(u + \sqrt{d}v)(x_1 - \sqrt{d}y_1)^n = 1 \, ,$$

da cui si ricava che:

$$u + \sqrt{d}v = x_n + \sqrt{d}y_n \ .$$

Pertanto, $u = x_n$ e $v = y_n$.

□

Osservazione 5.15. Si può dimostrare agevolmente che le soluzioni della equazione di Pell formano un gruppo rispetto alla "composizione" definita utilizzando le identità del Lemma 5.8 (cfr. anche Lemma 5.2, Corollario 5.9 ed Osservazione 5.10).
Inoltre, se $\mathbf{Sol}(d) := \{(x, y) \in \mathbb{Z} \times \mathbb{Z} : x^2 - dy^2 = 1\}$ è l'insieme di tutte le soluzioni dell'equazione di Pell, allora l'applicazione canonica:

$$\varphi : \mathbf{Sol}(d) \longrightarrow \mathbb{Z}[\sqrt{d}] := \{a + b\sqrt{d} : a, b \in \mathbb{Z}\} \ , \quad (x, y) \mapsto x + y\sqrt{d}$$

definisce un isomorfismo canonico tra $\mathbf{Sol}(d)$ ed il gruppo moltiplicativo $\mathrm{Im}(\varphi)$ (sottogruppo del gruppo moltiplicativo degli elementi non nulli di $\mathbb{Z}[\sqrt{d}]$) il quale è un gruppo moltiplicativo ciclico infinito generato da $\varepsilon_1 := x_1 + \sqrt{d}y_1$, dove (x_1, y_1) è la soluzione fondamentale dell'equazione di Pell (cfr. anche l'Esercizio 5.2).

5.17. Algoritmo elementare per determinare la soluzione fondamentale dell'equazione di Pell

Il Teorema precedente permette di trovare tutte le soluzioni dell'equazione di Pell non appena ne sia nota una non banale. Infatti, da una soluzione non banale è possibile, con un numero finito di passi, determinare la soluzione fondamentale. Un metodo (algoritmico elementare) per trovare effettivamente la soluzione fondamentale dell'equazione di Pell è il seguente:

Fissato d (intero positivo che non possiede fattori quadratici), si consideri la successione di interi positivi:

$$\{1 + dk^2 : k \geq 1\} \, .$$

Sia y_1 il più piccolo intero positivo tale che $1 + dy_1^2$ sia il quadrato di un intero positivo x_1. Allora (x_1, y_1) è la soluzione fondamentale dell'equazione di Pell.

Infatti, se (x, y) è una soluzione positiva dell'equazione di Pell, allora (per la scelta di y_1) $y \geq y_1$ e, quindi, $x = \sqrt{1 + dy^2} \geq \sqrt{1 + dy_1^2} = x_1$, da cui si ricava che $x + \sqrt{d}y \geq x_1 + \sqrt{d}y_1$.

Osservazione 5.16. Si noti che, in alcuni casi, cioè per valori particolari di d, è facile determinare la soluzione fondamentale (x_1, y_1) dall'equazione di Pell.

(i) Se $d = a^2 - 1$, con a intero ed $a > 1$, allora:

$$(x_1, y_1) = (a, 1) \ .$$

Infatti, $1 + k^2(a^2 - 1) = k^2a^2$ che è un quadrato per $k = 1$. Dunque $y_1 = 1$ e $x_1 = a$.

(ii) Se $d = a(a + 1)$, con a intero positivo, allora:

$$(x_1, y_1) = (2a + 1, 2) \ .$$

Infatti, è evidente che $(2a+1, 2)$ è una soluzione positiva dell'equazione di Pell. Inoltre, se esistesse una soluzione del tipo $(x, 1)$, cioè con $y = 1$, allora si avrebbe $x^2 = dy^2 + 1 = d + 1 = a^2 + a + 1 > a^2$, cioè $x > a$, ovvero $x \geq a+1$ e, quindi, $x^2 \geq a^2+2a+1 > a^2+a+1 = x^2$, donde un assurdo. Quindi, ogni altra soluzione positiva (x, y) deve avere $y \geq 2$ e, quindi, $x \geq 2a + 1$.

Segnaliamo anche che è possibile dimostrare, anche se ciò è meno semplice, che:

(iii) Se $d = a^2 + 1$, con a intero positivo, allora:

$$(x_1, y_1) = (2a^2 + 1, 2a) \ .$$

(iv) Se $d = a^2 + 2$, con a intero positivo, allora:

$$(x_1, y_1) = (a^2 + 1, a) \ .$$

Terminiamo il paragrafo mostrando come le soluzioni dell'equazione diofantea di Pell forniscono un metodo di approssimazione dei numeri reali che sono radici quadrate di interi positivi. Tale metodo fornisce un semplice esempio di uno strumento particolarmente importante per lo studio delle equazioni diofantee, quale è quello della *teoria delle approssimazioni diofantee*. Questa è una branca tecnica e specializzata della teoria dei numeri che ha avuto inizio verso la metà del XVIII secolo con il tentativo di risoluzione numerica delle equazioni diofantee, osservando la rapidità di convergenza di valori approssimati delle radici.

Proposizione 5.17. (L. Euler, 1759) *Se (x, y) è una soluzione positiva della equazione di Pell* (5.1)*, allora il numero razionale x/y approssima il numero reale $\sqrt{d}$ con un'accuratezza superiore all'inverso del quadrato del denominatore, cioè:*

$$x/y - \sqrt{d} < 1/y^2 \ .$$

Dimostrazione. $1 = x^2 - dy^2 = (x + y\sqrt{d})(x - y\sqrt{d})$ da cui:

$$(x - y\sqrt{d})/y = 1/[y(x + y\sqrt{d})] < 1/y^2\sqrt{d} < 1/y^2 \ .$$

□

Segnaliamo infine che, facendo uso della teoria delle "frazioni continue", si può dare un metodo effettivo, sistematico e generale per la risoluzione dell'equazione diofantea di Pell. Tale metodo, risalente a Wallis e Brouncker, consiste essenzialmente nel produrre approssimazioni razionali vieppiú precise di $\sqrt{d}$. Per maggiori dettagli rinviamo al volume di Allenby e Redfern [1, Section 11.6] oppure a quello di Hardy e Wright [6, p. 129–153 e p. 210]; cfr. anche Davenport [3, Ch. 4, Par. 11], Olds [10].

Osservazione 5.18. L'equazione lineare e l'equazione quadratica di Pell godono di una particolare proprietà tra le equazioni diofantee in due indeterminate: esse sono essenzialmente le uniche equazioni diofantee in due indeterminate che ammettono infinite soluzioni intere o quasi–intere (cioè, razionali con denominatore fissato). Questo fatto è una delle conseguenze della introduzione di metodi della geometria algebrica in aritmetica.

Precisamente, a partire dall'inizio del 1900 con i lavori di H. Poincaré la teoria delle equazioni diofantee si è sviluppata in modo più sistematico ed organico interpretando "geometricamente" problematiche di origine aritmetica. Le equazioni polinomiali in due indeterminate a coefficienti in $\mathbb{Z}(\subset \mathbb{Q})$,

da un punto di vista geometrico, definiscono delle curve algebriche nel piano. Pertanto, le equazioni diofantee di tale tipo possono essere classificate utilizzando un invariante geometrico: il "genere" g delle curve algebriche associate.

Le curve di genere $g = 0$ possono essere ricondotte (tramite trasformazioni birazionali) a rette o a coniche (cfr. anche l'Esercizio 5.8). Le equazioni di tali curve hanno, in generale, infinite soluzioni intere e, quindi, le curve hanno infiniti punti a coordinate razionali.

Le curve di genere $g = 1$ sono le "curve ellittiche", che possono essere ricondotte ad un'equazione cubica standard. Mordell nel 1922 ha dimostrato che un'equazione diofantea la cui curva associata ha genere 1 ha soltanto un numero finito di soluzioni in $\mathbb{Z}$, ma può avere infinite soluzioni razionali, "generate" però da un numero finito tra esse.

Mordell ha poi congetturato che ogni curva algebrica di genere $g > 1$ può avere soltanto un numero finito di punti razionali. Tale congettura è stata dimostrata da G. Faltings nel 1983, con un lavoro che gli è valso la Medaglia Fields.

5 Esercizi e complementi

5.1. Mostrare che la soluzione fondamentale dell'equazione di Pell è data da:

$$\begin{array}{ll} (3,2), & \text{se } d=2\,; \\ (2,1), & \text{se } d=3\,; \\ (9,4), & \text{se } d=5\,; \\ (5,2), & \text{se } d=6\,; \\ (8,3), & \text{se } d=7\,; \\ (3,1), & \text{se } d=8\,; \\ (19,6), & \text{se } d=10\,; \\ (10,3), & \text{se } d=11\,; \\ (7,2), & \text{se } d=12\,. \end{array}$$

Per $d = 13$, la soluzione fondamentale è data da $(649, 180)$ e per $d = 61$, poi, la soluzione fondamentale è data da $(1766319049, 226153980)$ e risulta ovviamente non "agevole" il suo calcolo, procedendo con il metodo elementare precedentemente descritto.

[*Suggerimento*. Utilizzare il semplice algoritmo descritto in 5.15]

5.2. Sia d un intero positivo che non è un quadrato. Supponiamo che l'equazione diofantea:

$$X^2 - dY^2 = -1 \tag{5.2.1}$$

ammetta almeno una soluzione. (Si osservi che in generale un'equazione del tipo (5.2.1) non è risolubile: ad esempio $X^2 - 3Y^2 = -1$ non è risolubile perché non è risolubile la congruenza $x^2 \equiv -1 \pmod 3$). Allora:

(a) Esiste sempre una soluzione positiva (u_1, v_1) di (5.2.1) per la quale $\gamma_1 := u_1 + v_1\sqrt{d} > 1$ è minimo. Tale soluzione è detta *soluzione fondamentale*.

(b) Siano (w_1, z_1) e (w_2, z_2) due soluzioni di (5.2.1). Siano $\alpha_1 := w_1 + z_1\sqrt{d}$ e $\alpha_2 := w_2 + z_2\sqrt{d}$. Verificare che $\alpha_1\alpha_2 = (w_1w_2 + dz_1z_2) + (w_1z_2 + w_2z_1)\sqrt{d}$ e che $(w_3 := w_1w_2 + dz_1z_2,\ z_3 := w_1z_2 + w_2z_1)$ è una soluzione dell'equazione di Pell (5.1).

(c) Se $\varepsilon_1 = \gamma_1^2 = u_2 + v_2\sqrt{d}$, allora (u_2, v_2) è la soluzione fondamentale della equazione di Pell (5.1), cioè tutte le soluzioni non banali di (5.1) sono date da $(\pm u_{2n}, \pm v_{2n})$ per $n \geq 1$, dove $\gamma_{2n} := \varepsilon_1^n =: u_{2n} + v_{2n}\sqrt{d}$, e per ogni scelta possibile del segno.

(d) Tutte e sole le soluzioni di (5.2.1) sono date da $(\pm u_{2n+1}, \pm v_{2n+1})$ con $n \geq 0$ e per ogni possibile scelta del segno.

[*Suggerimento*. Sia (u, v) una soluzione di (5.2.1) e sia $\gamma := u + \sqrt{d}v$.

(a) Come nel caso dell'equazione di Pell, $u > 0$ e $v > 0$ equivale a $\gamma > 1$. Il ragionamento che segue la Definizione 5.13. si può applicare pure in questo caso,

per cui è possibile trovare una soluzione positiva (u_1, v_1) di (5.2.1), in modo tale che $\gamma_1 := u_1 + \sqrt{d}v_1 > 1$ sia minimo.

(b) È conseguenza del Lemma 5.2.

(c) Per definizione della soluzione fondamentale ε_1 dell'equazione di Pell (5.1) (Definizione 5.13) e per il fatto che γ_1^2 è anch'essa soluzione di (5.1), si ha:

$$1 < \varepsilon_1 \leq \gamma_1^2 \ .$$

Se poniamo $\overline{\gamma}_1 := u_1 - \sqrt{d}v_1$, allora $\gamma_1\overline{\gamma}_1 = -1$, dunque:

$$-\overline{\gamma}_1 < -\overline{\gamma}_1\varepsilon_1 \leq \gamma_1 \ .$$

È subito visto che, se scriviamo $-\overline{\gamma}_1\varepsilon_1 = a+\sqrt{d}b$ (rispettivamente, $\gamma_1\varepsilon_1 = a'+\sqrt{d}b'$), allora (a, b) (rispettivamente, (a', b')) è una soluzione di (5.2.1). Dunque

$$-\overline{\gamma}_1 < -\overline{\gamma}_1\varepsilon_1 < 1 \quad \text{oppure} \quad 1 < -\overline{\gamma}_1\varepsilon_1 \leq \gamma_1 \ .$$

La prima eventualità è esclusa perché essa è equivalente a

$$1 < \gamma_1\varepsilon_1 < \gamma_1 \, ,$$

che è esclusa per la proprietà di minimalità di γ_1 tra le soluzioni positive di (5.2.1). Pertanto, $1 < -\overline{\gamma}_1\varepsilon_1 \leq \gamma_1$ e quindi $-\overline{\gamma}_1\varepsilon_1 = \gamma_1$ (sempre per la minimalità di γ_1), cioè $\varepsilon_1 = \gamma_1^2$.

(d) Se (a, b) è una soluzione positiva di (5.2.1) e se $\delta := a + \sqrt{d}b$, allora si può trovare un intero positivo $n \geq 1$ tale che

$$1 \leq \delta\gamma_1^{-n} < \gamma_1 = \varepsilon_1^2 \, ,$$

dunque

$$\varepsilon_1^{-1} \leq \delta\varepsilon_1^{-1}\gamma_1^{-n} < \varepsilon_1 \ .$$

Dove scrivendo $\delta\varepsilon_1^{-1}\gamma_1^{-n} = x + \sqrt{d}y$ si ha che (x, y) è una soluzione dell'equazione di Pell (5.1).

D'altro lato $1 < \varepsilon_1 < \gamma_1 = \varepsilon_1^2$, quindi $1 > \varepsilon_1^{-1} > \gamma_1^{-1}$, pertanto:

$$\gamma_1^{-1} \leq \delta\varepsilon_1^{-1}\gamma_1^{-n} < \gamma_1$$

da cui si ricava che $\delta\varepsilon_1^{-1}\gamma_1^{-n} = 1$ cioè $\delta = \varepsilon_1^{2n+1}$.]

5.3. Mostrare che l'equazione (5.2.1) *non* è risolubile se $d = p$ è un numero primo, con $p \equiv 3 \pmod 4$.

[*Suggerimento*. Se (5.2.1) è risolubile, allora la congruenza $X^2 \equiv -1 \pmod p$ è risolubile].

5.4. Sia d un intero positivo che non è un quadrato. Si consideri l'equazione diofantea:

$$X^2 - dY^2 = 4 \ . \tag{5.4.1}$$

(a) Mostrare che (5.4.1) è sempre risolubile, provando che se (x', y') è una soluzione dell'equazione di Pell (5.1), allora $(2x', 2y')$ è una soluzione di (5.4.1).

(b) Mostrare che *non* ogni soluzione di (5.4.1) è del tipo descritto in (a) (ad esempio: $(3, 1)$ per $d = 5$).

(c) Mostrare che esiste sempre una soluzione positiva minima di (5.4.1) (ξ_1, η_1) (cioè tale che $\zeta_1 := \xi_1 + \sqrt{d}\eta_1$ è minimo positivo), detta *soluzione fondamentale* di (5.4.1).

(d) Provare che tutte e sole le soluzioni non banali (a, b) di (5.4.1) sono tali che, se $\alpha := a + b\sqrt{d}$, allora

$$(\star_n) \qquad \frac{\alpha}{2} = \pm \left(\frac{\zeta_1}{2} \right)^n \quad \text{per qualche } n \geq 1 \ .$$

Viceversa se α verifica $(\star_n)$ per un qualche n, allora (a, b) è una soluzione di (5.4.1).

[*Suggerimento.* (a) si verifica in modo diretto.

(b) $3^2 - 5 \cdot 1^2 = 4$, però $3 + \sqrt{5} \neq 2x' + 2y'\sqrt{5}$ presi comunque $x', y' \in \mathbb{Z}$.

(c) Si noti che se (x, y) è una soluzione di (5.4.1), allora x e y sono entrambi pari oppure entrambi dispari. Pertanto, possiamo scrivere $x + y\sqrt{d} \equiv a(1 + \sqrt{d}) \pmod 2$, con $a = 0$ oppure $a = 1$. Diremo che (x, y) è una *soluzione dispari* se $a = 1$. Si noti che, se esiste una soluzione dispari di (5.4.1), allora necessariamente d deve essere dispari. In ogni caso, se $\alpha := x + y\sqrt{d}$ e $\beta := x' + y'\sqrt{d} \equiv b(1 + \sqrt{d}) \pmod 2$ sono due soluzioni di (5.4.1), allora $\alpha\beta \equiv ab(d + 1 + 2\sqrt{d}) \equiv 0 \pmod 2$, perché se $ab = 1$ allora d è dispari e quindi $d + 1$ è pari. In ogni caso, per ogni coppia di soluzioni (x, y), (x', y') di (5.4.1), si ha:

$$\frac{\alpha\beta}{2} = 2 \cdot \frac{\alpha}{2} \cdot \frac{\beta}{2} = u + v\sqrt{d} \quad \text{con } u, v \in \mathbb{Z} \quad \text{e} \quad u^2 - v^2 d = 4 \ .$$

Da questa osservazione discende la prima parte di (c).]

La seconda affermazione di (c) e (d) si dimostrano con un ragionamento analogo a quello del Teorema 5.14.

5.5. Sia d un intero positivo che non è un quadrato. Si consideri l'equazione diofantea:

$$X^2 - dY^2 = -4 \ . \tag{5.5.1}$$

(a) Mostrare che (5.5.1) non è sempre risolubile.

Se (5.5.1) è risolubile, allora mostrare che:

(b) Esiste una soluzione positiva minima (μ_1, ν_1) di (5.5.1) (cioè, tale che $\lambda_1 := \mu_1 + \sqrt{d}\nu_1$ è minimo positivo) detta *soluzione fondamentale*.

(c) Se $\zeta_1 := \xi_1 + \sqrt{d}\eta_1$, dove (ξ_1, η_1) è la soluzione fondamentale di (5.4.1), allora mostrare che:

$$\frac{\zeta_1}{2} = \left(\frac{\lambda_1}{2}\right)^2 .$$

(d) Siano $a, b \in \mathbb{Z}$ e sia $\alpha := a + \sqrt{d}b$. Mostrare che:

$$\frac{\alpha}{2} = \pm\frac{\lambda_1}{2}\left(\frac{\zeta_1}{2}\right)^n , \quad \text{con } n \in \mathbb{Z} ,$$

se e soltanto se (a, b) è una soluzione di (5.5.1).

[*Suggerimento*. Le dimostrazioni delle affermazioni sopra enunciate sono simili a quelle dell'Esercizio 5.2.]

5.6. Sia d un intero positivo che non è un quadrato. Si consideri l'equazione diofantea:

$$X^2 - dY^2 = n , \quad \text{con } n > 0 . \tag{5.6.1}$$

(a) Mostrare che (5.6.1) non è sempre risolubile.

(b) Mostrare che se (5.6.1) è risolubile, allora essa ammette infinite soluzioni. Più precisamente, mostrare se (a, b) è una soluzione di (5.6.1) e se $(\pm x_k, \pm y_k)$ è una soluzione dell'equazione di Pell (5.1), allora $(\pm ax_k \pm dby_k, \pm ay_k \pm bx_k)$, con una appropriata evidente scelta dei segni, è ancora una soluzione di (5.6.1). L'insieme costituito da tali soluzioni con $k \geq 1$ e da (a, b) è detto *classe di soluzioni* individuata da (a, b).

(c) Mostrare che l'equazione $X^2 - 2Y^2 = 49$ ammette come soluzioni $(7, 0)$ e $(9, 4)$ e che tali soluzioni individuano classi di soluzioni disgiunte.

Osservazione. Si può dimostrare che se (5.6.1) è risolubile allora, nella classe individuata da una qualunque soluzione, esiste una soluzione (a, b) tale che

$$\sqrt{n} < a \leq \sqrt{\Delta n} \quad \text{dove } \Delta := \frac{1}{2}\left(1 + \frac{x_1 + \sqrt{d}y_1}{(x_1 - 1) + y_1\sqrt{d}}x_1\right)$$

essendo (x_1, y_1) la soluzione fondamentale dell'equazione di Pell (5.1). Pertanto, da ciò si ricava facilmente che (5.6.1) ammette soltanto un numero finito di classi distinte di soluzioni. Per maggiori dettagli cfr. ad esempio LeVeque [8, Theorem 8.9].

[*Suggerimento*. (a) Ad esempio, se $d = 3$ ed $n = 2$ l'equazione diofantea $X^2 - 3Y^2 = 2$ non è risolubile, altrimenti sarebbe risolubile la congruenza $X^2 \equiv 2 \pmod 3$.

(b) Discende facilmente dal Lemma 5.2.

(c) Si noti che la soluzione fondamentale dell'equazione di Pell $X^2 - 2Y^2 = 1$ è data da $(3, 2)$. Allora, basta osservare che, se $\varepsilon_1 := 3 + 2\sqrt{2}$, si ha:

$$7 = (7 + 0\sqrt{2}) \neq (9 + 4\sqrt{2})\varepsilon_1^k , \quad \text{per ogni } k \geq 1. \,]$$

5.7. Nella stessa situazione dell'Esercizio 5.6, si consideri l'equazione diofantea:

$$X^2 - dY^2 = -n \ . \tag{5.7.1}$$

Mostrare che, *mutatis–mutandis*, valgono ancora gli enunciati (a), (b), (c) ed un enunciato "analogo" a quello dell'Osservazione dell'Esercizio 5.6. Più precisamente, nell'enunciato dell'"analogo" di quello dell'Osservazione precedente, si deve porre la disuguaglianza:

$$\sqrt{n} < a < \sqrt{\Delta' n} \ , \quad \text{dove } \Delta' := \frac{(x_1 + y_1\sqrt{d})x_1}{2(x_1 + 1 + y_1\sqrt{d})} \ .$$

[*Suggerimento*. Per (a) si vedano anche gli Esercizi 5.2 e 5.5 e per (b) l'Esercizio 5.6 (b). (c) Si noti che se (x, y) è una soluzione (5.7.1) e (u, v) è una soluzione di (5.2.1) allora, per il Lemma 5.2, $(xy \mp dyv, xv \mp yu)$ sono ancora soluzioni di (5.7.1). Si concluda ragionando in modo analogo a quello dell'Esercizio 5.6 (c), osservando che la soluzione fondamentale dell'equazione diofantea $X^2 - 2Y^2 = -1$ è determinata $\gamma_1 := 1 + \sqrt{2}$, cioè è data da $(1, 1)$ (e quindi, come è noto, $\gamma_1^2 = \varepsilon_1 = 3 + 2\sqrt{2}$ determina la soluzione fondamentale dell'Equazione di Pell $X^2 - 2Y^2 = 1$).]

5.8. Sia

$$aX^2 + bXY + cY^2 + dX + eY + f = 0 \tag{5.8.1}$$

la generale equazione diofantea quadratica in due indeterminate (a coefficienti in $\mathbb{Z}$). Mostrare che il calcolo delle soluzioni di tale equazione può essere ricondotto al calcolo delle soluzioni di una equazione di Pell, nelle indeterminate U e V, del tipo:

$$U^2 - sV^2 = t \ ,$$

dove $s := 4(b^2 - 4ac)$, $t := (2bd - 4ac)^2 - 4(b^2 - 4ac)(d^2 - af)$.

[*Suggerimento*. Si osservi che un'equazione quadratica:

$$AX^2 + BX + C = 0$$

con $A, B, C \in \mathbb{Q}$, $A \neq 0$ ha soluzioni in $\mathbb{Q}$ se, e soltanto se, $B^2 - 4AC = D^2$, per un qualche $D \in \mathbb{Q}$. Si noti, poi, che se $A, B, C \in \mathbb{Z}$, allora tale equazione ha soluzioni intere se, e soltanto se, $2A \mid (-B \pm D)$.

Si consideri ora la generale equazione quadratica a coefficienti in $\mathbb{Z}$ in due indeterminate (5.8.1). Ponendo $Y = y_0$ per un qualche $y_0 \in \mathbb{Z}$, l'equazione (5.8.1) nella sola indeterminata X, ha soluzioni razionali se, e soltanto se,

$$(by_0 + d)^2 - 4a(cy_0^2 + ey_0 + f) = (b^2 - 4c)y_0^2 + (2bd - 4ae)y_0 + d^2 - 4af = v^2$$

per un qualche $v \in \mathbb{Q}$, (si noti che dopo la sostituzione si ha: $A := a$, $B := by_0 + d$, $C := cy_0^2 + ey_0 + f$).

Lasciamo al Lettore il compito di esprimere esplicitamente, in questo caso, quando tali soluzioni sono intere.
Poniamo $p := b^2 - 4ac$ e supponiamo $p \neq 0$ (per evitare casi banali), $q := 2bd - 4ae$ e $r := d^2 - 4af$. Allora, l'equazione nell'indeterminata Y a coefficienti interi:

$$pY^2 + qY + (r - v^2) = 0$$

ha soluzioni razionali se, e soltanto se,

$$q^2 - 4p(r - v^2) = u^2 \,, \quad \text{per un qualche } u \in \mathbb{Q} \,.$$

Anche in questo caso, lasciamo al Lettore il compito di esprimere quando tali soluzioni sono intere.
Dunque, ci siamo ricondotti a considerare l'equazione di Pell in due indeterminate U e V del tipo:

$$U^2 - 4pV^2 = q^2 - 4pr \,.$$

Quindi, in definitiva, il saper risolvere tale equazione diofantea di Pell permette di saper risolvere l'equazione diofantea generale quadratica in due incognite (5.8.1).]

References

[1] R.B.J.T. Allenby, E.J. Redfern, *Introduction to Number Theory with Computing*, Arnold, 1989.

[2] E.T. Bell, "The Last Theorem", *MAA*, 1990.

[3] H. Davenport, *Aritmetica Superiore. Una introduzione alla Teoria dei Numeri*, Zanichelli, 1994.

[4] I.E. Dickson, *History of the theory of numbers*, (3 voll. 1920–1923) Ristampa Chelsea, New York 1974.

[5] H.M. Edwards, *Fermat's Last Theorem*, Springer, 1996 (new updated edition).

[6] G.H. Hardy, E.M. Wright, *An introduction to the theory of numbers*, Oxford 1954.

[7] E. Landau, *Elementary Number Theory*, Chelsea, New York 1958 (traduzione inglese di Vorlesungen über Zahlentheorie, vol. I, 1927).

[8] W.J. LeVeque, *Fundamentals of Numbers Theory*, Addison–Wesley, 1977.

[9] L.J. Mordell, *Diophantine equations*, Academic Press, 1969.

[10] C.D. Olds, *Continued fractions*, Random House, 1963.

[11] P. Ribenboim, *Fermat's Last Theorem for amateurs*, Springer, 1999

[12] H.E. Rose, *A course in number theory*, Oxford Science Pu., 1988.

[13] W. Sierpiński, *Elementary Theory of Numbers*, North–Holland 1988.

[14] S. Singh, *L'Ultimo Teorema di Fermat*, Rizzoli, 1997.

[15] A. van der Porten, *Notes on Fermat's Last Theorem*, Wiley, 1996.

[16] A. Weil, *Number Theory. An Approach through history*, Birkhäuser, 1984.

[17] E. Whitford, *The Pell's Equation*, New York 1912.