Jest to zapis odczytu wygłoszonego na VIII Festiwalu Nauki w Warszawie.

Skąd się wzięła liczba i

oraz nieliczne przykłady niecałkiem oczywistych jej zastosowań

Michał KRYCH, Warszawa

W tym tekście, jak wszędzie poza polskimi liceami, zbiór wszystkich liczb całkowitych oznaczany jest symbolem $\mathbb{Z}$

Definicja liczb zespolonych
Liczbami zespolonymi nazywamy liczby postaci $a+bi$, gdzie i oznacza jednostkę urojoną, przyjmujemy, że $i^2=-1$, oraz że a i b są liczbami rzeczywistymi. Suma liczb zespolonych $z_1=a+bi$ i $z_2=c+di$ to $z_1+z_2=(a+c)+(b+d)i$. Iloczyn liczb zespolonych $z_1=a+bi$ i $z_2=c+di$ to $z_1z_2=(ac-bd)+(ad+bc)i$. Zbiór wszystkich liczb zespolonych oznaczany jest (na całym świecie z wyjątkiem polskich szkół średnich) przez $\mathbb{C}$. ■

Stwierdzenie 1 (przemienność działań).
Dla dowolnych liczb zespolonych z_1, z_2 zachodzą równości

$$z_1+z_2=z_2+z_1 \qquad \text{oraz} \qquad z_1z_2=z_2z_1,$$

czyli dodawanie i mnożenie są działaniami przemiennymi.

Dowód. Uzasadniamy to w następujący sposób: $z_1+z_2=(a+c)+(b+d)i=$ $=(c+a)+(d+b)i=z_2+z_1$, bo wynik dodawania liczb rzeczywistych nie zależy od kolejności składników. Teraz mnożenie:

$$\begin{aligned} z_1z_2=(a+bi)(c+di)&=(ac-bd)+(ad+bc)i= \\ &=(ca-db)+(cb+da)i=(c+di)(a+bi)=z_2z_1, \end{aligned}$$

bo dodawanie i mnożenie liczb rzeczywistych są przemienne. ■

Zamiast pisać $a+0i$ będziemy pisać a, zamiast pisać $0+bi$ będziemy pisać bi. Liczby postaci bi, gdzie $b\in\mathbb{R}$, nazywać będziemy urojonymi. Dzięki tej umowie liczby rzeczywiste to szczególne liczby zespolone – „te w których nie ma i”. Liczbę a nazywamy częścią rzeczywistą liczby $z=a+bi$, piszemy $\operatorname{Re} z=a$; liczbę b – częścią urojoną liczby $z+a+bi$, piszemy $\operatorname{Im} z=b$.

W taki sam sposób sprawdzić można, że zachodzi

Stwierdzenie 2 (łączność, rozdzielność, istnienie różnicy i ilorazu).
Dla dowolnych liczb zespolonych z_1, z_2, z_3 zachodzą równosci
$(z_1+z_2)+z_3=z_1+(z_2+z_3)$ — dodawanie jest łączne,
$(z_1z_2)z_3=z_1(z_2z_3)$ — mnożenie jest łączne,
$z_1(z_2+z_3)=z_1z_2+z_1z_3$ — mnożenie jest rozdzielne względem dodawania.

Dla dowolnych liczb zespolonych z_1, z_2 istnieje dokładnie jedna liczba zespolona z taka, że $z_1+z=z_2$, liczba ta zwana jest różnicą liczb z_2 i z_1 i oznaczana symbolem z_2-z_1.

Dla dowolnych liczb zespolonych $z_1\neq 0$ i z_2 istnieje dokładnie jedna liczba zespolona z taka, że $z_1z=z_2$, liczba ta zwana jest ilorazem liczb z_2 i z_1 i oznaczana symbolem $\dfrac{z_2}{z_1}$ lub z_2/z_1.

Dowód. Jedynie dowód istnienia ilorazu różni się nieco od dowodu przemienności działań. Przyjmijmy, że $z_1=a+bi$, $z_2=c+di$. Szukamy liczby zespolonej $z=x+yi$, dla której $z_2=zz_1$, czyli $c+di=(a+bi)(x+yi)=$ $=(ax-by)+(ay+bx)i$. Ma więc być $c=ax-by$ i jednocześnie $d=ay+bx$. Otrzymaliśmy więc układ równań z niewiadomymi x, y. Mnożąc pierwsze z nich przez a, drugie przez b i dodając stronami otrzymujemy $ac+bd=(a^2+b^2)x$, zatem $x=\dfrac{ac+bd}{a^2+b^2}$, dzielenie jest wykonalne, bo $0\neq a+bi$, więc co najmniej jedna z liczb a, b jest różna od 0. Analogicznie otrzymujemy wzór $y=\dfrac{ad-bc}{a^2+b^2}$. ■

Wykazaliśmy wszystkie podstawowe własności działań. Jest oczywiste, że w zbiorze liczb zespolonych zachodzą równości $1\cdot z=z$, $0\cdot z=0$ i $0+z=z$ dla dowolnego $z\in\mathbb{C}$.

Na liczbach zespolonych możemy więc wykonywać działania tak, jak na liczbach rzeczywistych. Na przykład:

$$\frac{c+di}{a+bi} = \frac{(c+di)(a-bi)}{(a+bi)(a-bi)} = \frac{(c+di)(a-bi)}{a^2-(bi)^2} = \frac{(ac+bd)+(ad-bc)i}{a^2-b^2i^2} =$$
$$= \frac{(ac+bd)+(ad-bc)i}{a^2-b^2(-1)} = \frac{ac+bd}{a^2+b^2} + \frac{ad-bc}{a^2+b^2}i.$$

Niestety, nie wszystko jest tak jak w przypadku liczb rzeczywistych. W zbiorze $\mathbb{C}$ nie można w sensowny sposób wprowadzić nierówności. Nadamy temu zdaniu postać twierdzenia, a następnie udowodnimy je.

Twierdzenie o nieistnieniu nierówności w zbiorze liczb zespolonych
W zbiorze $\mathbb{C}$ nie istnieje relacja $\prec$ taka, że

1. Jeśli $z_1, z_2 \in \mathbb{C}$, to zachodzi dokładnie jedna z trzech możliwości: $\mathbf{z_1 = z_2}$ albo $z_1 \prec z_2$ albo $z_2 \prec z_1$ (każde dwie liczby można porównać);
2. jeśli $z_1 \prec z_2$ i $z_2 \prec z_3$, to $z_1 \prec z_3$ (nierówność ma być przechodnia);
3. jeśli $z_1 \prec z_2$ i $z \in \mathbb{C}$, to $z_1 + z \prec z_2 + z$ (do obu stron nierówności wolno dodać dowolną liczbę $z \in \mathbb{C}$);
4. jeśli $z_1 \prec z_2$ i $0 \prec z$, to $zz_1 \prec zz_2$ (nierówność wolno pomnożyć obustronnie przez dowolną liczbę z większą od 0).

Dowód. Załóżmy bowiem, że udało nam się w jakiś sposób zdefiniować nierówność $\prec$ w taki sposób, że spełnione są warunki 1 – 4. Jeśli $0 \prec z$, to $0 = 0 \cdot z \prec z \cdot z = z^2$, czyli kwadraty liczb dodatnich są dodatnie. Mamy oczywiście $z^2 = (-z)^2$. Jeśli $z \prec 0$, to $0 = z + (-z) \prec 0 + (-z) = -z$, zatem $0 \prec (-z)^2 = z^2$, więc również w tym przypadku $0 \prec z^2$. Wobec tego kwadraty liczb różnych od zera muszą być dodatnie. Mamy $1^2 = 1$ i $i^2 = -1$, zatem $0 \prec 1$ i jednocześnie $0 \prec -1$. Dodając do obu stron pierwszej z tych nierówności liczbę -1 otrzymujemy $-1 \prec (-1) + 1 = 0$, co przeczy temu, że $0 \prec -1$. Dowód został zakończony. ■

Okazało się więc, że liczb zespolonych porównywać się nie da. Można oczywiście definiować jakieś nierówności między liczbami zespolonymi rezygnując z części warunków 1 – 4, ale takie nierówności nie są użyteczne, więc na ogół nikt tego nie robi.

Liczby zespolone można traktować jako punkty płaszczyzny. Przyjmujemy, że część rzeczywista liczby zespolonej to pierwsza współrzędna (czyli pozioma), a część urojona to druga współrzędna (pionowa) punktu płaszczyzny. Przy takiej interpretacji suma $z_1 + z_2$ liczb zespolonych może być potraktowana jako koniec wektora, który jest sumą wektorów $\overrightarrow{0z_1}$ i $\overrightarrow{0z_2}$.

Definicja. Wartością bezwzględną $|z|$ liczby zespolonej $z = a + bi$ nazywamy liczbę $\sqrt{a^2+b^2}$. Argumentem $\operatorname{Arg} z$ liczby $z = a + bi \neq 0$ nazywamy dowolną liczbę φ taką, że $\cos\varphi = \dfrac{a}{\sqrt{a^2+b^2}}$ oraz $\sin\varphi = \dfrac{b}{\sqrt{a^2+b^2}}$. ■

Z definicji tej wynika, że $|z|$ to odległość punktu z od punktu 0 a argument liczby z, to kąt między wektorami $\overrightarrow{01}$ i $\overrightarrow{0z}$ mierzony w kierunku przeciwnym do ruchu wskazówek zegara.

Na przykład $\operatorname{Arg} 2 = 0$ lub $\operatorname{Arg} 2 = 2004\pi$, $\operatorname{Arg} i = \dfrac{\pi}{2}$ lub $\operatorname{Arg} i = -\dfrac{3\pi}{2}$,
$\operatorname{Arg}(-1+i) = \pi - \dfrac{\pi}{4} = \dfrac{3}{4}\pi$ itp. $|2| = 2 = |-2| = |2i| = |-2i|$,
$|1+i| = |-1+i| = |1-i| = |-1-i| = \sqrt{2}$ itp.

Nierówność trójkąta
Dla dowolnych liczb zespolonych z_1, z_2 zachodzi nierówność $|z_1 + z_2 \leqslant |z_1| + |z_2|$, jest ona równością jedynie wtedy, gdy punkty płaszczyzny odpowiadające liczbom $0, z_1, z_2$ leżą na jednej prostej, przy czym 0 między z_1 i z_2 (nieostro, jedna z liczb z_1, z_2 może być zerem). ■

Dowodu nie podajemy, bo wynika on ze znanych własności figur geometrycznych (np. trójkąta), a ci którzy ich nie pamiętają, mogą bez kłopotu wykazać,

że dla dowolnych liczb rzeczywistych a_1, b_1, a_2, b_2 zachodzi nierówność $\sqrt{(a_1+a_2)^2+(b_1+b_2)^2} \leqslant \sqrt{a_1^2+b_1^2}+\sqrt{a_2^2+b_2^2}$, staje się ona równością wtedy i tylko wtedy, gdy istnieje liczba rzeczywista $t \geqslant 0$ taka, że $z_1 = tz_2$ lub $z_2 = tz_1$.

Z równości $z = a+bi$, $r = |z|$, $\cos\varphi = \dfrac{a}{\sqrt{a^2+b^2}}$ i $\sin\varphi = \dfrac{b}{\sqrt{a^2+b^2}}$ wynika, że $z = r(\cos\varphi + i\sin\varphi)$. Zapisaliśmy liczbę z w *postaci trygonometrycznej.*

Załóżmy, że $z_1 = r_1(\cos\varphi_1 + i\sin\varphi_1)$ i $z_2 = r_2(\cos\varphi_2 + i\sin\varphi_2)$. Wtedy

$$z_1z_2 = r_1r_2\big(\cos\varphi_1\cos\varphi_2 - \sin\varphi_1\sin\varphi_2 + i(\cos\varphi_1\sin\varphi_2 + \cos\varphi_2\sin\varphi_1)\big) =$$
$$= r_1r_2\big(\cos(\varphi_1+\varphi_2) + i\sin(\varphi_1+\varphi_2)\big)$$

– skorzystaliśmy tu z wzorów $\cos(\varphi_1+\varphi_2) = \cos\varphi_1\cos\varphi_2 - \sin\varphi_1\sin\varphi_2$ oraz $\sin(\varphi_1+\varphi_2) = \cos\varphi_1\sin\varphi_2 + \cos\varphi_2\sin\varphi_1$, z którymi wiele osób spotyka się czasem w szkołach. Wykazaliśmy w ten sposób, że wartość bezwzględna iloczynu dwu liczb zespolonych równa jest iloczynowi ich wartości bezwzględnych, a argument iloczynu dwu liczb zespolonych równa jest sumie ich argumentów. Stosując otrzymany wzór wielokrotnie otrzymujemy

Wzór de Moivre'a

$$\big(r(\cos\varphi + i\sin\varphi)\big)^n = r^n\big(\cos(n\varphi) + i\sin(n\varphi)\big). \blacksquare$$

Z tego wzoru wynika, że dla każdej liczby zespolonej $w \neq 0$ i każdej liczby naturalnej n istnieje dokładnie n różnych liczb zespolonych z_1, z_2,…,z_n takich, że $z_j^n = w$ dla $j = 1, 2, \ldots, n$. Załóżmy bowiem, że $w = \varrho(\cos\psi + i\sin\psi)$. Jeśli $z = r(\cos\varphi + i\sin\varphi)$ i $w = z^n$, to muszą być spełnione równości $\varrho = r^n$ oraz $n\varphi = \psi + 2k\pi$ dla pewnej liczby całkowitej k. Wynika stąd, że $r = \sqrt[n]{\varrho}$, a więc r jest wyznaczone jednoznacznie. Musi też być $\varphi = \dfrac{\psi}{n} + \dfrac{2k\pi}{n}$. Zastępując liczbę k liczbą $k+n$ zwiększamy kąt φ o 2π, co nie zmienia liczby z. Różne liczby z otrzymujemy przyjmując kolejno $k = 0$, $k = 1$,…, $k = n-1$. Otrzymujemy więc dokładnie n różnych wartości. Łatwo zauważyć, że odpowiadające im punkty płaszczyzny są wierzchołkami n–kąta foremnego wpisanego w okrąg o promieniu $r = \sqrt[n]{\varrho}$. Jeśli $w = 1$, to wśród tych liczb jest liczba 1.

Definicja pierwiastka algebraicznego z liczby zespolonej

Algebraicznym pierwiastkiem n–tego stopnia z liczby zespolonej w nazywamy każdą liczbę zespoloną z, dla której $w = z^n$. ■

Przykłady

1. Pierwiastkami algebraicznymi stopnia 2 z liczby $1 = \cos 0 + i\sin 0$ są $z_1 = \cos\dfrac{0\pi}{2} + i\sin\dfrac{0\pi}{2} = \cos 0 + i\sin 0 = 1$ oraz $z_2 = \cos\dfrac{2\pi}{2} + i\sin\dfrac{2\pi}{2} = \cos\pi + i\sin\pi = -1$.
2. Pierwiastkami algebraicznymi stopnia 3 z liczby $1 = \cos 0 + i\sin 0$ są $z_1 = \cos\dfrac{0\pi}{2} + i\sin\dfrac{0\pi}{2} = 1$, $z_2 = \cos\dfrac{2\pi}{3} + i\sin\dfrac{2\pi}{3} = -\dfrac{1}{2} + i\dfrac{\sqrt{3}}{2}$ oraz $z_3 = \cos\dfrac{4\pi}{2} + i\sin\dfrac{4\pi}{2} = -\dfrac{1}{2} - i\dfrac{\sqrt{3}}{2}$.
3. Pierwiastkami algebraicznymi stopnia 3 z liczby $-1 = \cos\pi + i\sin\pi$ są $z_1 = \cos\dfrac{\pi}{3} + i\sin\dfrac{\pi}{3} = \dfrac{1}{2} + i\dfrac{\sqrt{3}}{2}$, $z_2 = \cos\dfrac{\pi+2\pi}{3} + i\sin\dfrac{\pi+2\pi}{3} = -1$ oraz $z_3 = \cos\dfrac{\pi+4\pi}{3} + i\sin\dfrac{\pi+4\pi}{3} = \dfrac{1}{2} - i\dfrac{\sqrt{3}}{2}$.
4. Ponieważ $\cos 2\alpha + i\sin 2\alpha = (\cos\alpha + i\sin\alpha)^2 = cos^2\alpha + 2i\cos\alpha\sin\alpha + i^2\sin^2\alpha = \cos^2\alpha - \sin^2\alpha + 2i\cos\alpha\sin\alpha$, części rzeczywiste są równe i części urojone są równe, więc $\cos 2\alpha = \cos^2\alpha - \sin^2\alpha$ i $\sin 2\alpha = 2\sin\alpha\cos\alpha$.
5. Ponieważ $\cos 3\alpha + i\sin 3\alpha = (\cos\alpha + i\sin\alpha)^3 = cos^3\alpha + 3i\cos^2\alpha\sin\alpha + 3i^2\cos\alpha\sin^2\alpha + i^3\sin^3\alpha = \cos^3\alpha - 3\cos\alpha\sin^2\alpha + i\big(3\cos^2\alpha\sin\alpha - \sin^3\alpha\big)$, więc $\cos 3\alpha = \cos^3\alpha - 3\cos\alpha\sin^2\alpha = 4\cos^3\alpha - 3\cos\alpha$ oraz $\sin 3\alpha = 3\cos^2\alpha\sin\alpha - \sin^3\alpha = 3\sin\alpha - 4\sin^3\alpha$.

Widzimy więc, że za pomocą liczb zespolonych można powiązać wzory na $\cos n\alpha$ i $\sin n\alpha$ z dwumianem Newtona i można przestać poszukiwać tych wzorów w tablicach.

Definicja sprzężenia.
Jeśli $z = a + bi$, gdzie $a, b \in \mathbb{R}$, to liczbę $\overline{z} = a - bi$ nazywamy sprzężoną do liczby z. ■

Na przykład $\overline{2 - 3i} = 2 + 3i$, $\overline{13} = 13$, $\overline{i} = -i$.

Liczba z jest rzeczywista wtedy i tylko wtedy, gdy , $z = \overline{z}$. Jeśli $z \notin \mathbb{R}$, to $\overline{z} \in \mathbb{C}$ jest jedyną liczbą taką, że $z + \overline{z} \in \mathbb{R}$ i jednocześnie $z \cdot \overline{z} \in \mathbb{R}$. Prosty dowód tego stwierdzenia pozostawiam Czytelnikom w charakterze ćwiczenia. Mamy też

$$z \cdot \overline{z} = (a + bi)(a - bi) = a^2 + b^2 = |z|^2, \ z + \overline{z} = 2 \operatorname{Re} z \text{ oraz } z - \overline{z} = 2i \operatorname{Im} z.$$

Możemy więc napisać $\operatorname{Re} z = \frac{1}{2}(z + \overline{z})$ i $\operatorname{Im} z = \frac{1}{2i}(z - \overline{z})$. Punkty płaszczyzny odpowiadające liczbom z i $\overline{z}$ są symetryczne względem osi rzeczywistej.

Przypomnijmy, że argument iloczynu dwu liczb zespolonych równy jest sumie argumentów składników. Jest to własność przypominająca nieco logarytm (logarytm iloczynu to suma logarytmów czynników). Wykorzystując tę analogię można w sposób sensowny zdefiniować potęgę o wykładniku zespolonym i podstawie, którą matematycy uważają za najważniejszą, tzn. o podstawie e.

Liczby zespolone są używane, bo w niektórych sytuacjach nie sposób się bez nich obejść. Historycznie pierwszym przypadkiem tego rodzaju był chyba wzór na pierwiastki równania trzeciego stopnia:

$$\text{jeżeli } x^3 + px + q = 0, \text{ to } x = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{p^3}{27} + \frac{q^2}{4}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{p^3}{27} + \frac{q^2}{4}}}.$$

Wyprowadzenie tego wzoru nie jest długie. Załóżmy, że $u + v = x$. Ma więc być spełniona równość $0 = (u + v)^3 + p(u + v) + q = u^3 + v^3 + q + (u + v)(p + 3uv)$. Wystarczyłoby więc znaleźć parę liczb rzeczywistych u, v taką, że $u^3 + v^3 = -q$ i jednocześnie $uv = -\frac{p}{3}$. W dziedzinie rzeczywistej równanie $uv = -\frac{p}{3}$ jest równoważne równaniu $u^3 v^3 = -\frac{p^3}{27}$. Bez trudu stwierdzamy więc, że liczby u^3, v^3 muszą być pierwiastkami równania kwadratowego $t^2 + qt - \frac{p^3}{27} = 0$. Muszą więc być spełnione równości $u^3 = -\frac{q}{2} - \sqrt{\frac{p^3}{27}}$ i $v^3 = -\frac{q}{2} + \sqrt{\frac{p^3}{27}}$ (lub odwrotnie, ale przecież u, v grają w naszym rozumowaniu te same role). Stąd otrzymujemy równość

$$x = u + v = \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{p^3}{27} + \frac{q^2}{4}}} + \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{p^3}{27} + \frac{q^2}{4}}}$$

Pokażemy, że stosowanie tego wzoru może być kłopotliwe. Niech $p = -63$, $q = -162$, zajmujemy się więc równaniem $x^3 - 63x - 162 = 0$. Mamy $\frac{p^3}{27} + \frac{q^2}{4} = \left(\frac{p}{3}\right)^3 + \left(\frac{q}{2}\right)^2 = (-21)^3 + 81^2 = -2700 < 0$. Teraz z tej liczby należy wyciągnąć pierwiastek kwadratowy. Ten pierwiastek nie jest liczbą rzeczywistą! Można pomyśleć, że to dlatego, że nasze równanie nie ma rozwiązań rzeczywistych. Tak jednak nie jest. Mamy bowiem $(-3)^3 - 63 \cdot (-3) - 162 = 0$, $(-6)^3 - 63 \cdot (-6) - 162 = 0$ oraz $9^3 - 63 \cdot 9 - 162 = 0$, więc nasze równanie ma trzy pierwiastki rzeczywiste. Otrzymujemy więc wzory

$$-3 = \sqrt[3]{81 + \sqrt{-2700}} + \sqrt[3]{81 - \sqrt{-2700}},$$
$$-6 = \sqrt[3]{81 + \sqrt{-2700}} + \sqrt[3]{81 - \sqrt{-2700}},$$
$$9 = \sqrt[3]{81 + \sqrt{-2700}} + \sqrt[3]{81 - \sqrt{-2700}}.$$

Wygląda to nieco podejrzanie: prawe strony są równe, a lewe różne. To jednak tylko pozór. Są dwie wartości pierwiastka kwadratowego z danej liczby zespolonej różnej od 0 i trzy wartości pierwiastka trzeciego stopnia. Przy tej interpretacji można się spodziewać do trzydziestu sześciu pierwiastków tego równania. To jednak nie jest możliwe. Równanie stopnia trzeciego ma najwyżej

trzy pierwiastki (po prostu nie można wybierać wartości tych pierwiastków w sposób dowolny). Udowodniono, że nie jest możliwe napisanie wzorów na pierwiastki równania stopnia trzeciego z użyciem dodawania, odejmowania, mnożenia, dzielenia i pierwiastkowania, które nie prowadziłyby do wyciągania pierwiastków kwadratowych z liczb ujemnych w przypadku rzeczywistych współczynników i trzech rzeczywistych pierwiastków! Oznacza to, że w tym przypadku bez liczb zespolonych obyć się nie można. Zaczęto ich więc używać w XVI wieku, choć „ich nie było". Zostały ostatecznie zaakceptowane na początku XIX wieku, gdy C.F.Gauss pokazał, że można je potraktować jako punkty płaszczyzny i że wtedy działania na liczbach zespolonych zaczynają mieć sens geometryczny. Dziś trudno sobie wyobrazić matematykę bez ich użycia.

Gaussowi udało się również podać poprawny dowód zasadniczego twierdzenia algebry (nazwa dziś stosowana). Udowodnił on mianowicie, że każdy wielomian stopnia co najmniej 1 o współczynnikach zespolonych ma co najmniej jeden pierwiastek zespolony. Choć udowodnione twierdzenie sformułowane zostało w terminach algebraicznych, to dowód Gaussa był w swej istocie geometryczny (czy jak byśmy dziś powiedzieli — topologiczny, tej nazwy w czasach autora dowodu jeszcze nie używano). Sformułujemy teraz to twierdzenie wraz z pewnym wnioskiem.

Zasadnicze twierdzenie algebry.
Jeśli $a_0, a_1, \ldots, a_n \in \mathbb{C}$ oraz $n \geqslant 1$ i $a_n \neq 0$, to istnieje co najmniej jedna liczba zespolona z_1 taka, że $a_0 + a_1 z_1 + \cdots + a_n z_1^n = 0$, czyli: *każdy wielomian stopnia większego od 0 o współczynnikach zespolonych ma co najmniej jeden pierwiastek zespolony.* ■

Dowód tego twierdzenia nie mieści się w tym wykładzie. Z twierdzenia tego wynika, że jeśli $p(z) = a_0 + a_1 z + \cdots + a_n z^n$, $n \geqslant 1$, $a_n \neq 0$, to istnieje co najmniej jedna liczba zespolona z_1 taka, że dla pewnych liczb zespolonych $b_0, b_1, \ldots, b_{n-1}$ wzór $p(z) = (z - z_1)(b_0 + b_1 z + \cdots + b_{n-1} z^{n-1})$ zachodzi dla każdej liczby zespolonej z (twierdzenie Bézout). Oczywiście $b_{n-1} = a_n \neq 0$. Jeśli $n - 1 \geqslant 1$, to istnieje liczba z_2 taka, że $b_0 + b_1 z_2 + \cdots + b_{n-1} z_2^{n-1} = 0$, więc powtarzając rozumowanie stwierdzamy istnienie liczb $c_0, c_1, \ldots, c_{n-2}$ takich, że dla każdej liczby zespolonej z zachodzi równość $p(z) = (z - z_1)(z - z_2)(c_0 + c_1 z + \cdots + c_{n-2} z^{n-2})$. Tę zabawę można kontynuować dopóki nie rozłożymy wielomianu p na iloczyn wielomianów stopnia pierwszego i stałej: $p(z) = a_n(z - z_1)(z - z_2) \cdot \ldots \cdot (z - z_n)$. Wywnioskowaliśmy właśnie z zasadniczego twierdzenia algebry

Wniosek. Każdy wielomian o współczynnikach zespolonych możemy przedstawić w postaci iloczynu wielomianów stopnia pierwszego o współczynnikach zespolonych. ■

Wniosek ten może być zastosowany również do wielomianów, których współczynnikami są liczby rzeczywiste, w końcu liczby rzeczywiste są również liczbami zespolonymi (bardzo szczególnymi). Takie wielomiany będziemy nazywać *rzeczywistymi*. Wtedy z tego wniosku można wywnioskować nieco więcej.

Twierdzenie o nierzeczywistych pierwiastkach wielomianu rzeczywistego.
Jeśli $a_0, a_1, \ldots, a_n \in \mathbb{R}$, $n \geqslant 1$ i $a_n \neq 0$ oraz $a_0 + a_1 z + a_2 z_2 + \cdots + a_n z^n = 0$, to również $a_0 + a_1 \bar{z} + a_2 \bar{z}^2 + \cdots + a_n \bar{z}^n = 0$, tzn. jeśli liczba zespolona z jest pierwiastkiem wielomianu o współczynnikach rzeczywistych to jej sprzężenie $\bar{z}$ również jest pierwiastkiem tego wielomianu.

Dowód. Mamy $0 = \bar{0} = \overline{a_0 + a_1 z + a_2 z_2 + \cdots + a_n z^n} = \bar{a}_0 + \bar{a}_1 \bar{z} + \bar{a}_2 \bar{z}^2 + \cdots + \bar{a}_n \bar{z}^n = a_0 + a_1 \bar{z} + a_2 \bar{z}_2 + \cdots + a_n \bar{z}^n$ — trzecia równość wynika z własności sprzężenia, czwarta — z tego, że współczynniki $a_0, a_1, \ldots, a_n$ są rzeczywiste. Dowód został zakończony. ■

Widzimy więc, że nierzeczywiste pierwiastki wielomianu rzeczywistego występują parami. Jeśli z_1, jest nierzeczywistym pierwiastkiem wielomianu rzeczywistego

$p(z)$, to również liczba $z_2 = \bar{z}_1$ jest jego pierwiastkiem, a ponieważ $z_1 \neq \bar{z}_1 = z_2$, więc wielomian $p(z)$ jest podzielny przez wielomian $(z - z_1)(z - z_2) =$
$= z^2 - (z_1 + z_2)z + z_1 z_2 = z^2 - (z_1 + \bar{z}_1)z + z_1\bar{z}_1 = z^2 - 2\operatorname{Re} z_1 z + |z_1|^2$.
Współczynniki tego ostatniego wielomianu są liczbami rzeczywistymi!
Oczywiście ten wielomian kwadratowy nie ma pierwiastków rzeczywistych (bo ma nierzeczywiste, a ma ich tylko 2 jako wielomian stopnia drugiego). Stąd łatwo już wnioskujemy, że

Twierdzenie o rozkładzie wielomianu rzeczywistego na czynniki nierozkładalne.
Każdy wielomian rzeczywisty stopnia nie mniejszego niż 1 można przedstawić w postaci iloczynu wielomianów rzeczywistych stopnia pierwszego i drugiego o ujemnych wyróżnikach. ■

Okazało się więc, że przynajmniej z punktu widzenia rozwiązywania równań wielomianowych dalsze rozszerzanie zapasu liczb nie jest potrzebne. (Nie jest też w pewnym sensie możliwe, ale wyjaśnienie odpowiedniego twierdzenia zajęłoby za dużo miejsca.)

Ze znanego wzoru na sumę pierwszych n wyrazów ciągu geometrycznego wyprowadzimy wzór na sumy: $\sin\varphi + \sin(2\varphi) + \cdots + \sin n\varphi$ oraz
$\cos\varphi + \cos(2\varphi) + \cdots + \cos(n\varphi)$.

Mamy

$$\begin{aligned}
&\cos\varphi + \cos(2\varphi) + \cdots + \cos(n\varphi) + i\sin\varphi + i\sin(2\varphi) + \cdots + i\sin n\varphi = \\
&\quad = \cos\varphi + i\sin\varphi + \cos(2\varphi) + i\sin(2\varphi) + \cdots + \cos(n\varphi) + i\sin(n\varphi) = \\
&\quad = \cos\varphi + i\sin\varphi + (\cos\varphi + i\sin\varphi)^2 + \cdots + (\cos\varphi + i\sin\varphi)^n \overset{q=\cos\varphi+i\sin\varphi}{=\!=\!=\!=\!=\!=\!=\!=\!=\!=\!=} \\
&\quad = q + q^2 + \cdots + q^n = \\
&\quad = q\frac{q^n - 1}{q - 1} = (\cos\varphi + i\sin\varphi)\frac{\cos(n\varphi) + i\sin(n\varphi) - 1}{\cos\varphi + i\sin\varphi - 1} = \\
&\quad = (\cos\varphi + i\sin\varphi)(\cos\varphi - i\sin\varphi - 1)\frac{\cos(n\varphi) + i\sin(n\varphi) - 1}{(\cos\varphi + i\sin\varphi - 1)(\cos\varphi - i\sin\varphi - 1)} = \\
&\quad = \big(1 - (\cos\varphi + i\sin\varphi)\big)\frac{\cos(n\varphi) + i\sin(n\varphi) - 1}{(\cos\varphi - 1)^2 + \sin^2\varphi} = \\
&\quad = \frac{\cos(n\varphi) + i\sin(n\varphi) - 1 - \cos(n+1)\varphi - i\sin(n+1)\varphi + \cos\varphi + i\sin\varphi}{2(1 - \cos\varphi)}.
\end{aligned}$$

Ponieważ części rzeczywiste równych liczb zespolonych są równe, więc

$$\begin{aligned}
&\cos\varphi + \cos(2\varphi) + \cdots + \cos(n\varphi) = \\
&= \operatorname{Re}\Big(\frac{\cos(n\varphi) + i\sin(n\varphi) - 1 - \cos(n+1)\varphi - i\sin(n+1)\varphi + \cos\varphi + i\sin\varphi}{2(1 - \cos\varphi)}\Big) = \\
&= \frac{\cos(n\varphi) - 1 - \cos(n+1)\varphi + \cos\varphi}{2(1 - \cos\varphi)} = \frac{2\sin\frac{\varphi}{2}\sin\frac{(2n+1)\varphi}{2} - 2\sin^2\frac{\varphi}{2}}{4\sin^2\frac{\varphi}{2}} = \\
&= \frac{\sin\frac{(2n+1)\varphi}{2} - \sin\frac{\varphi}{2}}{2\sin\frac{\varphi}{2}} = \frac{\sin\frac{n\varphi}{2}\cos\frac{(n+1)\varphi}{2}}{\sin\frac{\varphi}{2}}
\end{aligned}$$

— zastosowaliśmy w końcówce wzory trygonometryczne znane kiedyś licelistom.
Porównując części urojone otrzymujemy wzór:

$$\sin\varphi + \sin(2\varphi) + \cdots + \sin n\varphi = \frac{\sin\frac{n\varphi}{2}\sin\frac{(n+1)\varphi}{2}}{\sin\frac{\varphi}{2}}.$$

Zapewne wiele osób dowodziło za pomocą indukcji matematycznej uzyskane wzory, ale nam udało się je uzyskać jako wniosek z wzoru na sumę wyrazów skończonego ciągu geometrycznego. Żadna hipoteza na wstępie nie była potrzebna! Również w tym przypadku pokazaliśmy rozwiązanie problemu, w którego sformułowaniu liczb zespolonych nie ma, natomiast pojawiają się w rozwiązaniu.

Znacznie bardziej spektakularne było rozumowanie Eulera z połowy XVIII w uzasadniające szczególny przypadek twierdzenia Fermata. Udowodnił on mianowicie, że równanie $x^3 + y^3 = z^3$ nie rozwiązań całkowitych poza takimi, w których jedna z niewiadomych jest równa 0. Spróbujemy opisać to rozumowanie.

Autor tego tekstu nie widział dowodu twierdzenia Fermata w przypadku $n = 3$ nie korzystającego z liczb zespolonych.

Niech $\omega = -\frac{1}{2} + i\frac{\sqrt{3}}{2} = \cos 120^\circ + i \sin 120^\circ$, więc $\omega^3 = 1$, zatem $0 = \omega^3 - 1 = (\omega - 1)(\omega^2 + \omega + 1)$. Wobec tego $\omega^2 + \omega + 1 = 0$, co można też zapisać w postaci $\omega^2 = -\omega - 1$. Będziemy rozważać zbiór $\mathbb{Z}[\omega]$, którego elementami są liczby zespolone postaci $a + b\omega$, gdzie a i b oznaczają liczby całkowite. Zauważmy po pierwsze, że suma, różnica i iloczyn liczb z $\mathbb{Z}[\omega]$ są również w $\mathbb{Z}[\omega]$. W przypadku sumy i różnicy jest to zupełnie oczywiste. Sprawdzimy, że jest tak również w przypadku iloczynu. Niech a, b, c, d będą liczbami całkowitymi. Mamy $(a + b\omega)(c + d\omega) = ac + (bc + ad)\omega + bd\omega^2 = ac + (bc + ad)\omega + bd(-1 - \omega) = ac - bd + (bc + ad - bd)\omega$. Udało się.

Jeśli liczby a, b, c, d są całkowite i $a + b\omega = c + d\omega$, to $a = c$ i $b = d$. DLACZEGO?

Niech $N(z) = |z^2| = z\bar{z}$ dla $z \in \mathbb{Z}[\omega]$. Mamy więc

$$N(a + b\omega) = (a + b\omega)(a + b\bar{\omega}) = a^2 + ab(\omega + \bar{\omega}) + b^2\omega\bar{\omega} = a^2 - ab + b^2$$

— skorzystaliśmy z oczywistych równości $\omega + \bar{\omega} = -\frac{1}{2}$ oraz $\omega\bar{\omega} = 1$. Ponieważ $|z_1 z_2| = |z_1| \cdot |z_2|$ dla dowolnych liczb zespolonych z_1, z_2, więc $N(z_1 z_2) = N(z_1)N(z_2)$ dla dowolnych liczb $z_1, z_2 \in \mathbb{Z}[\omega]$.

Załóżmy teraz, że dla pewnych liczb $z_1, z_2 \in \mathbb{Z}[\omega]$ zachodzi wzór $z_1 z_2 = 1$. Niech $z_1 = a + b\omega$, $z_2 = c + d\omega$, $a, b, c, d \in \mathbb{Z}$. Mamy więc $1 = N(z_1 z_2) = N(z_1)N(z_2) = (a^2 - ab + b^2)(c^2 - cd + d^2)$. Wobec tego $1 = a^2 - ab + b^2 = \frac{1}{2}\big(a^2 + b^2 + (a - b)^2\big)$, czyli $2 = a^2 + b^2 + (a - b)^2$ i analogicznie $2 = c^2 + d^2 + (c - d)^2$. Suma trzech kwadratów liczb całkowitych równa jest 2, zatem dwa z tych kwadratów są równe 1, a trzeci równy jest 0. Stąd wnioskujemy, że musi zachodzić jeden z warunków: $a = b = 1$; $a = b = -1$; $a = \pm 1, b = 0$; $a = 0, b = \pm 1$. Oznacza to, że jeśli $z_1 \in \mathbb{Z}[\omega]$ jest taką liczbą, że $\frac{1}{z_1} \in \mathbb{Z}[\omega]$, to $z_1 = \pm(1 + \omega) = \mp\omega^2 = \mp\bar{\omega}$ (bo $1 + \omega = -\omega^2$ i $\omega^2 \cdot \omega = 1 = \bar{\omega} \cdot \omega$) lub $z_1 = \pm 1$ lub $z_1 = \pm\omega$. Takie liczby nazywamy dzielnikami jedynki w zbiorze $\mathbb{Z}[\omega]$.

W zbiorze $\mathbb{Z}[\omega]$ można zajmować się teorią podzielności. Wyjaśnimy pokrótce jak można to robić. Mówimy, że liczba $z_1 \in \mathbb{Z}[\omega]$ jest dzielnikiem liczby $z_2 \in \mathbb{Z}[\omega]$ wtedy i tylko wtedy, gdy $\frac{z_2}{z_1} \in \mathbb{Z}[\omega]$. Mówimy, że liczba $z \in \mathbb{Z}[\omega]$ jest odwracalna wtedy i tylko wtedy, gdy $\frac{1}{z} \in \mathbb{Z}[\omega]$. Mówimy, że liczba $z \in \mathbb{Z}[\omega]$ jest *rozkładalna* (to odpowiada liczbie całkowitej złożonej) wtedy i tylko wtedy, gdy $z = z_1 z_2$ dla pewnych *nieodwracalnych* liczb $z_1, z_2 \in \mathbb{Z}[\omega]$. Liczby *nierozkładalne* (czyli pierwsze) to te, które są jednocześnie nierozkładalne i nieodwracalne. W poprzednim akapicie wykazaliśmy, że jedynymi odwracalnymi liczbami w $\mathbb{Z}[\omega]$ są ± 1, $\pm\omega$ oraz $\pm\omega^2$. Liczba $2 + \omega$ jest nierozkładalna, bo z równości $2 + \omega = z_1 z_2$ wynika, że $3 = 2^2 - 2 \cdot 1 + 1^2 = N(2 + \omega) = N(z_1)N(z_2)$, a stąd wynika, że $N(z_1) = 1$ i $N(z_2) = 3$ lub $N(z_1) = 3$ i $N(z_2) = 1$. W pierwszym przypadku liczba z_1 jest odwracalna, a w drugim — liczba z_2. Natomiast liczba 3 jest rozkładalna (czyli złożona!), bo $3 = (2 + \omega)(2 + \bar{\omega})$.

Można udowodnić, że każdą liczbę z $\mathbb{Z}[\omega]$ można przedstawić w postaci iloczynu liczb pierwszych w $\mathbb{Z}[\omega]$ i to na jeden sposób. To twierdzenie sformułujemy dokładnie po zakończeniu dowodu twierdzenia Fermata w przypadku $n = 3$, a potem podamy jego dowód. Twierdzenie to w szkołach jest stosowane w odniesieniu do zbioru liczb całkowitych, choć nie jest nawet formułowane. Kiedyś w ogóle nie widziano potrzeby jego dowodu. Potem okazało się, że rozpatrując podzielność w różnych zbiorach można się natknąć na takie, w których przedstawienie liczby w postaci iloczynu liczb pierwszych jest niejednoznaczne! To oznacza, że twierdzenie o jednoznaczności przedstawienia liczby w postaci iloczynu czynników pierwszych z pewnością wymaga

uzasadnienia. Nawet w przypadku liczb całkowitych uzasadnienie to nie jest na tyle proste, by mogło być omawiane w szkole.

Twierdzenie o dzieleniu z resztą.
Dla dowolnych liczb $w, z \in \mathbb{Z}[\omega]$, $z \neq 0$ istnieją liczby κ, ϱ takie, że $w = \kappa z + \varrho$ i $N(\varrho) < N(z)$, κ nazywamy ilorazem, a ϱ resztą z dzielenia liczby w przez liczbę z.

Nie twierdzimy, że iloraz i reszta są zdefiniowane jednoznacznie, bo tak nie jest.

Dowód. Niech a, b, c, d oznaczają liczby całkowite takie, że $w = a + b\omega$, $z = c + d\omega$. Niech r, s będą takimi liczbami wymiernymi, że $r + s\omega = \dfrac{w}{z} = \dfrac{a+b\omega}{c+d\omega} = \dfrac{(a+b\omega)(c+d\bar\omega)}{(c+d\omega)(c+d\bar\omega)} = \dfrac{ac+bd+cd\omega+\bar a d\omega}{c^2-cd+d^2}$. Niech m, n oznaczają liczby całkowite takie, że $|r-m| \leqslant \dfrac{1}{2}$ i $|s-n| \leqslant \dfrac{1}{2}$. Mamy więc

$$a + b\omega = (c+d\omega)(r+s\omega) = (c+d\omega)(m+n\omega) + (c+d\omega)\big((r-m)+(s-n)\omega\big).$$

Mamy $(c+d\omega)\big((r-m)+(s-n)\omega\big) = a + b\omega - (c+d\omega)(m+n\omega) \in \mathbb{Z}[\omega]$, bo a, b, c, d, m, n są liczbami całkowitymi. Mamy też

$$\begin{aligned}\big|(c+d\omega)\big((r-m)+(s-n)\omega\big)\big|^2 &= \\ &= (c^2-cd+d^2)\big((r-m)^2-(r-m)(s-n)+(s-n)^2\big) \leqslant \\ &\leqslant (c^2-cd+d^2)\big(\tfrac14+\tfrac14+\tfrac14\big) = \tfrac34(c^2-cd+d^2) = \tfrac34 N(z) < N(z).\end{aligned}$$

Wystarczy więc przyjąć $\kappa = m + n\omega$ i $\varrho = a + b\omega - (c+d\omega)(m+n\omega)$. ■

Będziemy wielokrotnie zajmować się podzielnością przez liczbę $\lambda := 1 - \omega = 2 + \omega^2 = 2 + \bar\omega$ lub jej potęgi. Pokazaliśmy już, że jest to liczba pierwsza w $\mathbb{Z}[\omega]$, oraz że jest ona dzielnikiem liczby 3. Ponieważ $N(\lambda) = 3$ i dla każdej liczby $z \in \mathbb{Z}[\omega]$ mamy $N(z) \neq 2$, bo nie istnieją liczby całkowite a, b, dla których $a^2 - ab + b^2 = 2$, więc resztami z dzielenia przez λ mogą być jedynie liczby $0, \pm1, \pm\omega, \pm\omega^2$. Ponieważ $\omega = (\omega - 1) + 1 = -\lambda + 1$, więc resztę ω możemy zastąpić resztą 1. Podobnie z tego, że $\omega^2 = \omega^2 - 1 + 1 = \lambda(\omega + 1) + 1$, wynika, że resztę ω^2 można zastąpić przez 1. Analogicznie zamiast $-\omega$ i $-\omega^2$ możemy mieć resztę -1. Wobec tego

można przyjąć, że jedynymi resztami z dzielenia przez λ są liczby $0, -1, 1$.

Zauważmy jeszcze, że

jeśli $\varepsilon_1, \varepsilon_2 \in \mathbb{Z}[\omega]$ i $\varepsilon_1 = 1 = \varepsilon_2$,
to liczba λ^2 nie jest dzielnikiem liczby $\varepsilon_1 + \varepsilon_2$.

Wynika to stąd, że dla każdej liczby $z \in \mathbb{Z}[\omega], z \neq 0$ zachodzi nierówność $|z| \geqslant 1$ oraz tego, że $|\lambda^2| = 3 > 2 \geqslant |\varepsilon_1 + \varepsilon_2|$.

Zajmiemy się równaniem $x^3 + y^3 = \mu z^3$, gdzie μ oznacza dowolny dzielnik jedności, czyli jedną z liczb $\pm1, \pm\omega, \pm\omega^2$. Wykażemy, że niezależnie od wyboru μ jedyne rozwiązanie tego równania w zbiorze $\mathbb{Z}[\omega]$ to takie, dla których $xyz = 0$, czyli trywialne. Gdy $\mu = 1$, równanie ma postać $x^3 + y^3 = z^3$.

Zauważmy najpierw, że jeśli λ jest dzielnikiem liczby $x - 1 \in \mathbb{Z}$, to λ^4 jest dzielnikiem liczby $x^3 - 1$. Jeśli bowiem λ jest dzielnikiem liczby $x - 1$, to istnieje liczba $t \in \mathbb{Z}$ taka, że $x - 1 = t\lambda$. Mamy wtedy

$$\begin{aligned}x^3 - 1 = t^3\lambda^3 + 3t^3\lambda^3 + 3t\lambda &= \lambda^3(t^3 - t) + 3\lambda^2t^2 + (\lambda^3 + 3\lambda)t = \\ &\overset{\lambda\bar\lambda=3}{=\!=\!=} \lambda^3(t^3 - t) + \lambda^3\bar\lambda t^2 + (\lambda^3 + \lambda^2\bar\lambda)t = \\ &\overset{\substack{\bar\lambda=-\bar\omega\lambda \\ \lambda+\bar\lambda=3=\lambda\bar\lambda}}{=\!=\!=\!=\!=} \lambda^3 t(t-1)(t+1) - \bar\omega\lambda^4t^3 - \bar\omega\lambda^4t.\end{aligned}$$

Drugi i trzeci składnik są podzielne przez λ^4; pierwszy też, bo jedna z liczb $t, t-1, t+1$ jest podzielna przez λ.

Załóżmy najpierw, że λ **nie** jest dzielnikiem xyz. Wykażemy, rozumując nie wprost, że nie zachodzi równość $x^3 + y^3 = \mu z^3$. Można przyjąć, że liczba $x - 1$ jest podzielna przez λ (jeśli nie, to trójkę x, y, z zastępujemy trójką $-x, -y, -z$.). Ponieważ λ nie jest dzielnikiem liczby y, więc λ jest dzielnikiem

dokładnie jednej z liczb $y-1$, $y+1$. W pierwszym przypadku λ^4 jest dzielnikiem liczby $x^3-1+y^3-1=\mu z^3-2$. Wobec tego jeśli λ jest dzielnikiem liczby $z-1$, to λ^4 jest dzielnikiem liczby z^3-1 i wobec tego również dzielnikiem liczby $\mu z^3-2-\mu(z^3-1)=-2+\mu$ wbrew temu, że $N(\lambda^4)=81>9\geqslant|-2+\mu|^2=$ $=N(-2+\mu)$. Jeśli λ jest dzielnikiem liczby $z+1$, jest wtedy dzielnikiem liczby $-z-1$, zatem λ^4 jest dzielnikiem liczby $(-z)^3-1=-z^3-1$, więc λ^4 jest dzielnikiem liczby z^3+1 i liczby $\mu(z^3+1)$. Wobec tego λ^4 jest dzielnikiem liczby $\mu z^3-2-\mu(z^3+1)=-2-\mu$, co jest niemożliwe (jak wyżej). Wykazaliśmy, jeśli λ jest dzielnikiem liczby $y-1$, to trójka x,y,z nie jest rozwiązaniem badanego równania. Załóżmy teraz, że λ jest dzielnikiem liczby $y+1$. Wtedy λ^4 jest dzielnikiem liczby y^3+1, zatem λ^4 jest dzielnikiem liczby $x^3-1+y^3+1=$ $=x^3+y^3=\mu z^3$, a ponieważ μ jest dzielnikiem jedności, więc λ^4 musi dzielić liczbę z^3, a założyliśmy, że tak nie jest. Wykazaliśmy zatem, że jeśli trójka x,y,z jest rozwiązaniem, to λ jest dzielnikiem co najmniej jednej z liczb x,y,z.

Możemy oczywiście zakładać, że każde dwie liczby wybrane z trójki x,y,z są względnie pierwsze. Jeśli nie są, to jakaś liczba pierwsza $p\in\mathbb{Z}[\omega]$ jest ich wspólnym dzielnikiem i wobec tego jest dzielnikiem trzeciej z nich. To pozwala podzielić obie strony równania przez p^3. Po skończonej liczbie kroków dochodzimy do trójki liczb parami względnie pierwszych. Dalej zakładamy, że liczba λ jest dzielnikiem dokładnie jednej z liczb x,y,z.

Załóżmy najpierw, że λ jest dzielnikiem z. Wobec tego nie jest dzielnikiem iloczynu xy. Ponieważ λ nie jest dzielnikiem liczby 2, więc reszty z dzielenia x^3 i y^3 przez λ muszą się „znieść”. Oznacza to, że jedna z niech jest równa 1, a druga jest równa -1. Załóżmy dla ustalenia uwagi, że λ jest dzielnikiem $x-1$ i $y+1$. Wtedy λ^4 jest dzielnikiem liczby $x^3-1+y^3+1=x^3+y^3=\mu z^3$, **zatem λ^2 jest dzielnikiem z.**

Mamy $\mu z^3=x^3+y^3=(x+y)(x+\omega y)(x+\omega^2 y)$. Ponieważ λ^2 jest dzielnikiem z, więc λ^6 jest dzielnikiem z^3. Stąd wynika, że co najmniej jedna z liczb $x+y$, $x+\omega y$, $x+\omega^2 y$ jest podzielna przez λ^2. Bez straty ogólności możemy założyć, że λ^2 jest dzielnikiem $x+y$ (jeśli trójka x,y,z jest rozwiązaniem naszego równania, to również trójki $x,\omega y,z$ i x,ω^2,z są rozwiązaniami tego równania.). Mamy $x+\omega y=x+y+(\omega-1)y=x+y-\lambda y$. Ponieważ liczba y jest niepodzielna przez λ a liczba $x+y$ jest podzielna przez λ^2, więc liczba $x+\omega y$ jest podzielna przez λ ale przez λ^2 już nie. To samo dotyczy liczby $x+\omega^2 y=$ $=x+y+(\omega^2-1)y\overset{\omega=1-\lambda}{=\!=\!=}x+y-\lambda(2-\lambda)y$.

Mamy $(x+y)-(x+\omega y)=\lambda y$ oraz $(x+\omega y)-\omega(x+y)=\lambda x$. Stąd i z tego, że liczby x,y są względnie pierwsze (w $\mathbb{Z}[\omega]$) wynika, że największym wspólnym dzielnikiem liczb $x+y$ i $x+\omega y$ jest λ. W taki sam sposób można wykazać, że λ jest największym wspólnym dzielnikiem $x+y$ i $x+\omega^2 y$ oraz $x+\omega y$ i $x+\omega^2 y$.

Stąd i z tego, że rozkład na czynniki pierwsze w $\mathbb{Z}[\omega]$ jest jednoznaczny wynika, że istnieją dzielniki jedności $\mu_1,\mu_2,\mu_3\in\mathbb{Z}[\omega]$ oraz parami względnie pierwsze, niepodzielne przez λ liczby $\alpha,\beta,\gamma\in\mathbb{Z}[\omega]$ i liczba naturalna $k\geqslant 2$ takie, że

$$x+y=\mu_1\alpha^3\lambda^k,\quad x+\omega y=\mu_2\beta^3\lambda\quad\text{i}\quad x+\omega^2 y=\mu_3\gamma^3\lambda.$$

Mnożymy stronami drugą z tych równości przez ω, trzecią przez ω^2, następnie dodajemy wszystkie trzy. Ponieważ $1+\omega+\omega^2=0$, więc otrzymujemy

$$0=\mu_1\alpha^3\lambda^k+\omega\mu_2\beta^3\lambda+\omega^2\mu_3\gamma^3\lambda.$$

Nie pozostaje nic lepszego do zrobienia niż podzielenie otrzymanej równości stronami przez $\mu_2\omega\lambda$ i skorzystanie z tego, że liczba $k+2$, czyli wykładnik z jakim wchodzi λ w rozkład na czynniki pierwsze liczby z^3, jest podzielna przez 3:

$$0=\frac{\mu_1}{\mu_2\omega}\alpha^3\lambda^{k-1}+\beta^3+\omega\frac{\mu_3}{\mu_2}\gamma^3,$$

przy czym liczba $k-1$ jest podzielna (w zbiorze $\mathbb{Z}$ tym razem) przez 3. Przyjmijmy $\varepsilon_2=-\dfrac{\mu_1}{\mu_2\omega}$, $\varepsilon_1=\omega\dfrac{\mu_3}{\mu_2}$, $x_1=\beta$, $y_1=\gamma$, $z_1=\alpha\lambda^{(k-1)/3}$. Mamy

$|\varepsilon_1| = |\varepsilon_2| = 1$, $\varepsilon_1, \varepsilon_2, x_1, y_1, z_1 \in \mathbb{Z}[\omega]$ oraz

$$x_1^3 + \varepsilon_1 y_1^3 = \varepsilon_2 z_1^3.$$

Liczby x_1, y_1 nie są podzielne przez λ, natomiast liczba z_1 jest podzielna przez λ, bo $k+2 \geqslant 6$. Tak jak poprzednio możemy przyjąć, że liczba $x_1 - 1$ dzieli się przez λ. Wynika stąd, że $x^3 - 1$ dzieli się przez λ^4, a wobec tego $\varepsilon_1 y^3 + 1 = \varepsilon_2 z_1^3 - (x_1^3 - 1)$ jest podzielne przez λ^2. Jeśli $y+1$ dzieli się przez λ, to $y^3 + 1$ dzieli się przez λ^4 i wobec tego $-\varepsilon_1 + 1 = \varepsilon_2 z_1^3 - (x_1^3 - 1) - \varepsilon_1(y^3 + 1)$ dzieli się przez λ^2. Jest to możliwe tylko wtedy, gdy $\varepsilon_1 = 1$. Jeśli λ jest dzielnikiem $y - 1$, to wynikiem analogicznego rozumowania jest równość $\varepsilon_1 = -1$. Mamy więc do czynienia albo z równaniem $x_1^3 + y_1^3 = \varepsilon_2 z_1^3$ albo z równaniem $x_1^3 + (-y_1)^3 = \varepsilon_2 z_1^3$. Jest to równanie takiej samej postaci jak równanie $x^3 + y^3 = z^3$ z tym, że liczba λ występuje w rozkładzie liczby z_1 z wykładnikiem $\dfrac{k-1}{3}$, więc o 1 mniejszym niż w rozkładzie liczby z. Powtarzając to rozumowanie wielokrotnie musimy dojść w końcu do równania tej samej postaci, w którym prawa strona podzielna przez λ już nie jest, a to jak wykazaliśmy wcześniej jest niemożliwe.

Pozostał do rozpatrzenia ostatni przypadek: λ jest dzielnikiem jednej z liczb x, y i nie jest dzielnikiem liczby z. Dla ustalenia uwagi załóżmy, że λ jest dzielnikiem liczby x i że liczby y, z są niepodzielne przez λ. Podobnie jak poprzednio możemy przyjąć, że liczba $y - 1$ jest podzielna przez λ. Stąd wynika, że liczba $x^3 + y^3 - 1 = \mu z^3 - 1 = \mu z(z^2 - 1) + \mu z - 1$ jest podzielna przez λ^3, zatem liczba $\mu z - 1$ również. Wobec tego albo liczba $\mu - 1$ albo liczba $-\mu - 1$ jest podzielna przez λ^3, więc również przez λ^2. Musi więc być $\mu = 1$ lub $\mu = -1$. Pozwala to na przepisanie równania $x^3 + y^3 = \mu z^3$ w postaci $(-y)^3 + z^3 = x^3$ lub w postaci $(-y)^3 + (-z)^3 = x^3$. W obu przypadkach prawa strona jest podzielna przez λ, a to jest w świetle poprzednich wyników jest niemożliwe.

Teraz pora na obiecane twierdzenie o jednoznaczności rozkładu. Zaczniemy od definicji największego wspólnego dzielnika dwu liczb. Ponieważ w zbiorze liczb zespolonych nie można wprowadzić sensownej nierówności, więc mamy drobny problem.

Twierdzenie o największym wspólnym dzielniku dwu liczb z $\mathbb{Z}[\omega]$.
Dla dowolnych dwu liczb $z_1, z_2 \in \mathbb{Z}[\omega]$, $z_1 \neq 0 \neq z_2$ istnieje liczba D, która jest dzielnikiem obu liczb z_1, z_2 taka, że jeśli liczba d jest dzielnikiem obu liczb z_1, z_2, to liczba d jest dzielnikiem liczby D.

Dowód. Niech A będzie zbiorem wszystkich liczb postaci $xz_1 + yz_2$, gdzie $x, y \in \mathbb{Z}[\omega]$. Niech $D = x_0 z_1 + y_0 z_2$ oznacza ten element zbioru A, którego norma $N(D) = D\bar{D}$ jest najmniejsza wśród norm dodatnich. Jasne jest, że jeśli d jest wspólnym dzielnikiem liczb z_1 i z_2, to jest dzielnikiem każdej liczby postaci $xz_1 + yz_2$, więc jest również dzielnikiem wybranej przez nas liczby D. Z twierdzenia o dzieleniu z resztą wynika, że istnieją takie liczby $q_1, r_1 \in \mathbb{Z}[\omega]$, że $z_1 = q_1 D + r_1$ i $N(r_1) < N(z_1)$ — podzieliliśmy liczbę z_1 z resztą przez liczbę D. Mamy $r_1 = z_1 - q_1 D = (1 - q_1 x_0) z_1 - q_1 y_0 z_2 \in A$. Ponieważ $0 \leqslant N(r_1) < N(D)$, więc $N(r_1) = 0$, czyli $z_1 = q_1 D$. W ten sposób wykazaliśmy, że liczba D jest dzielnikiem liczby z_1. Tak samo dowodzimy, że D jest dzielnikiem liczby z_2. ■

Definicja największego wspólnego dzielnika dwu liczb z $\mathbb{Z}[\omega]$.
Jeśli $z_1, z_2 \in \mathbb{Z}[\omega]$, $z_1 \neq 0 \neq z_2$, to największym wspólnym dzielnikiem liczb z_1, z_2 nazywamy taką liczbę $D \in \mathbb{Z}[\omega]$, która jest dzielnikiem obu liczb z_1, z_2, i jeśli liczba d jest dzielnikiem obu liczb z_1, z_2, to liczba d jest dzielnikiem liczby D. ■

Największy wspólny dzielnik nie jest wyznaczony jednoznacznie, bo pomnożywszy go przez dowolny dzielnik jedności otrzymujemy inny największy wspólny dzielnik. Bez trudu można stwierdzić, że jest to jedyna niejednoznaczność: jeśli D_1, D_2 są dwoma największymi wspólnymi dzielnikami liczb z_1, z_2, to istnieją liczby $\mu, \nu \in \mathbb{Z}[\omega]$ takie, że $D_1 = \mu D_2$ i $D_2 = \nu D_1$, zatem $D_1 = \mu\nu D_1$, więc $\mu\nu = 1$.

Lemat o liczbie pierwszej dzielącej iloczyn dwu liczb.
Jeśli $p \in \mathbb{Z}[\omega]$ jest liczbą pierwszą i jest dzielnikiem iloczynu $z_1 z_2$, to jest dzielnikiem co najmniej jednej z liczb z_1, z_2.

Dowód. Załóżmy, że p nie jest dzielnikiem liczby z_1. Ponieważ p jest liczbą pierwszą, więc największym wspólnym dzielnikiem liczb p i z_1 jest 1, zatem istnieją liczby $x, y \in \mathbb{Z}[\omega]$ takie, że $xz_1 + yp = 1$. Wobec tego $z_2 = xz_1z_2 + ypz_2$. Z założenia wynika, że liczba p jest dzielnikiem z_1z_2, zatem jest dzielnikiem obu składników prawej strony równości, więc jest dzielnikiem prawej strony, a to oznacza, że również lewej, czyli z_2. ■

Z definicji liczby złożonej (rozkładalnej) $z \in \mathbb{Z}[\omega]$ wynika, że jest ona iloczynem dwu liczb nieodwracalnych z_1, z_2, więc liczb o normach większych od 1. Mamy $N(z) = N(z_1)N(z_2)$, zatem $1 < N(z_1), N(z_2) < N(z)$. Wynika stąd od razu, że każdą liczbę nieodwracalną można przedstawić w postaci iloczynu liczb pierwszych.

Twierdzenie o jednoznaczności rozkładu na czynniki pierwsze.
Jeśli $z \in \mathbb{Z}[\omega]$ jest liczba nierozkładalną, to istnieją liczby pierwsze $p_1, p_2, \ldots, p_n \in \mathbb{Z}[\omega]$ takie, że $z = p_1p_2 \ldots p_n$. Jeśli liczby $q_1, q_2, \ldots, q_m \in \mathbb{Z}[\omega]$ są pierwsze i $z = q_1q_2 \ldots q_m$, to $m = n$ i po ewentualnej zmianie numeracji ilorazy $\frac{p_j}{q_j}$ są dzielnikami jedności w $\mathbb{Z}[\omega]$.

Dowód. Istnienie rozkładu wykazaliśmy przed twierdzeniem. Załóżmy, że liczba $z = p_1p_2 \ldots p_n$ ma dwa istotnie różne rozkłady na czynniki pierwsze $p_1, p_2, \ldots, p_n \in \mathbb{Z}[\omega]$. Ponadto każda liczba mająca rozkład na mniej niż n czynników pierwszych ma tylko jeden taki rozkład. Niech $z = q_1q_2 \ldots q_m$ i niech liczby $q_1, q_2, \ldots, q_m \in \mathbb{Z}[\omega]$ będą pierwsze. Ponieważ q_1 jest dzielnikiem iloczynu $p_1(p_2 \ldots p_n)$, więc jest dzielnikiem p_1 lub jest dzielnikiem iloczynu $p_2p_3 \ldots p_n$. Jeśli q_1 jest dzielnikiem p_1, to ponieważ obie te liczby są pierwsze, więc iloraz $\mu = \frac{p_1}{q_1}$ jest dzielnikiem jedności, więc równość $p_1p_2 \ldots p_n = q_1q_2 \ldots q_m$ można podzielić stronami przez q_1, co przeczy temu, że rozkład $p_1p_2 \ldots p_n$ był najkrótszy wśród niejednoznacznych. Wobec tego q_1 jest dzielnikiem $p_2p_3 \ldots p_n$. Powtarzając to rozumowanie jeszcze co najwyżej $n-2$ razy stwierdzamy, że któryś z ilorazów $\frac{p_j}{q_1}$ musi być liczbą odwracalną w $\mathbb{Z}[\omega]$. Stąd wynika, że równość $p_1p_2 \ldots p_n = q_1q_2 \ldots q_m$ można podzielić przez q_1, czyli skrócić rozkład niejednoznaczny jakoby najkrótszy. ■

Przykład niejednoznaczności rozkładu na czynniki pierwsze
W zbiorze $\mathbb{Z}[\sqrt{-5}]$, złożonym z liczb postaci $a + b\sqrt{-5} = a + bi\sqrt{5}$, gdzie a, b są liczbami całkowitymi, jednoznaczności rozkładu na czynniki pierwsze nie ma. Można, podobnie jak poprzednio zdefiniować normę: $N(a + b\sqrt{-5}) = (a + b\sqrt{-5})\overline{(a + b\sqrt{-5})} = a^2 + 5b^2$, jest ona oczywiście liczbą całkowitą dodatnią z wyjątkiem jednego przypadku: $a = b = 0$. Bez trudu stwierdzić można, że $N(z_1z_2) = |z_1z_2|^2 = |z_1|^2|z_2|^2 = N(z_1)N(z_2)$. Mamy $N(3) = 9$ i $N(2 + \sqrt{-5}) = 9$. Ponieważ nie istnieje liczba $z \in \mathbb{Z}[\sqrt{-5}]$ taka, że $N(z) = 3$ (równanie $a^2 + 5b^2 = 3$ nie ma rozwiązań w liczbach całkowitych), więc jeśli $z_1z_2 = 3$, to $N(z_1) = 1$ i $N(z_2) = 3$ (lub odwrotnie), ale to oznacza, że liczba z_1 jest dzielnikiem jedności. Wynika stąd, że liczba 3 jest pierwsza (w $\mathbb{Z}[\sqrt{-5}]$). To uzasadnienie pasuje również do liczby $2 + \sqrt{-5}$, więc ona też jest pierwsza. Również liczba $2 - \sqrt{-5}$ jest pierwsza. Jeśli $x + y\sqrt{-5} \in \mathbb{Z}[\sqrt{-5}]$ i $2 + \sqrt{-5} = 3(x + y\sqrt{-5})$, to $9 = N(2 + \sqrt{-5}) = N(3)N(x + y\sqrt{-5}) = 9(x^2 + 5y^2)$, zatem $x^2 + 5y^2 = 1$, więc $y = 0$ i $x = \pm 1$, ale nie jest prawdą, że $2 + \sqrt{-5} = \pm 3$. Wobec tego rozkłady $3 \cdot 3 = 9 = (2 + \sqrt{-5})(2 - \sqrt{-5})$ są istotnie różne! ■

Przygotowując ten tekst autor korzystał z książki K. Ireland, M. Rosen, *A Classical Introduction to Modern Number Theory*, Springer Verlag Inc., New York, 1982.