Musimy wiedzieć. Ale czy będziemy wiedzieć?

*Michał SKRZYPCZAK**

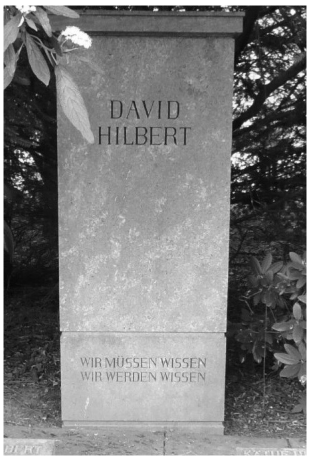
* Wydział Matematyki, Informatyki i Mechaniki, Uniwersytet Warszawski

Jest to tekst związany z odczytem wygłoszonym na LXI Szkole Matematyki Poglądowej pt. *Matematyczne zmiany*, Wola Ducka, luty 2020. Artykuł jest przedrukiem z czasopisma *Matematyka Poglądowa* (7), 2020.

Tytuł artykułu podejmuje polemikę z wypowiedzią Davida Hilberta wygłoszoną w Królewcu w 1930 roku:

Wir müssen wissen.
Wir werden wissen!

której tekst umieszczono później na grobie Hilberta.



Hilbert konsekwentnie głosił możliwość udzielenia przez matematykę odpowiedzi na każde pytanie. Wstęp do swoich 23 problemów (Paryż, 1900) zakończył:

In der Mathematik gibt es kein Ignorabimus.

Redakcja

Był pewien starożytny grecki geodeta, którego zadaniem było odmierzać prostokątne pastwiska na zboczach Olimpu. Zauważył on, że ilekroć pastwisko ma wymiary a na b, to długość jego przekątnej c spełnia zależność $c^2 = a^2 + b^2$. Jakkolwiek prawidłowość ta zawsze okazywała się prawdą, ów geodeta był przecież człowiekiem sumiennym i czuł się w obowiązku za każdym razem zmierzyć ową przekątną, zanim wpisał jej długość do Formularza Charakterystyki Gruntu (druk $P572$). Wszak nie ma żadnego powodu, by zależność prawdziwa w przypadku tysiąca innych pastwisk sprawdziła się w jakimś zupełnie nowym przypadku – byłaby to bezzasadna wiara w prawo serii!

Czy w takim razie geodetów należy dzielić na dwie kategorie: sumiennych, którzy sprawdzają wszystkie spostrzeżone zależności we wszystkich poszczególnych przypadkach, oraz lekkomyślnych, którzy bezpodstawnie wierzą w prawo serii? Otóż okazuje się, że jest też możliwość pośrednia, dostrzeżona przez Euklidesa. Można bowiem sformułować pewne minimalne wymagania (zwane *aksjomatami*), które winno spełniać pastwisko. Aksjomaty te to powinny być prawdy proste i oczywiste – a najlepiej jeszcze takie, których prawdziwość w przypadku nowo obmierzanego pastwiska łatwo sprawdzić. Na bazie tych aksjomatów można następnie, metodą rozumowania matematycznego, wydedukować rozmaite konsekwencje, w tym powyższą zależność $c^2 = a^2 + b^2$. I tak, przystępując do pomiarów w nowym miejscu, sprawdziwszy wpierw, że spełnia ono aksjomaty, niejako za darmo dostajemy gwarancję, że spełnia ono też wszystkie zależności z tych aksjomatów wyprowadzone.

Powyższa przenośnia ma obrazować metodę postępowania, na jakiej bazuje cała współczesna matematyka. Bo jakkolwiek jej celem jest często badanie konkretnych obiektów (pastwisk), to jednak nie jest to badanie empiryczne, w którym pieczołowicie sprawdzamy nasze zależności – zwykle zresztą jest to niemożliwe, bo większość prawidłowości wymagałaby sprawdzenia nieskończenie wielu przypadków. Zamiast tego formułujemy zbiory aksjomatów opisujących badany obiekt. W zależności od sytuacji aksjomaty te przyjmujemy „na wiarę” lub dowodzimy w oparciu o jakąś szerszą teorię. Wtedy reszta pracy matematycznej staje się działaniem ścisłym i formalnym: szukamy dowodu, że interesująca nas prawidłowość jest konsekwencją przyjętych aksjomatów. Pytanie tylko, czy taki dowód zawsze istnieje, nawet jeśli prawidłowość jest w jakimś sensie *prawdą*?

Zanim pójdziemy dalej w tych rozważaniach, musimy wprowadzić trochę oznaczeń. Konkretne obiekty matematyczne (pastwiska) będziemy nazywali *modelami* i oznaczali symbolami $\mathcal{M}$, $\mathcal{M}'$. Przykładami takich modeli są liczby rzeczywiste $\mathbb{R}$ z operacjami dodawania i mnożenia, oznaczane $\langle\mathbb{R},(+),(\cdot)\rangle$, czy liczby naturalne $\mathbb{N}$ z tymi samymi operacjami $\langle\mathbb{N},(+),(\cdot)\rangle$. Pojedyncze prawidłowości matematyczne wyrażające pewne własności modeli (jak na przykład warunek $(a+b)^2 = a^2+2ab+b^2$) będziemy nazywali *formułami* i oznaczali φ, ψ. I tak przemienność dodawania wyraża formuła $\varphi \equiv (\forall x, y.\ x+y = y+x)$ mówiąca, że dla każdej pary liczb x i y ich suma jest przemienna. *Aksjomatyka* lub *teoria* to nic innego jak pewien zbiór formuł $\Gamma = \{\varphi_1, \ldots\}$.

Rozważmy konkretny model $\mathcal{M}$ i formułę φ. Jeżeli φ jest *prawdziwa* w modelu $\mathcal{M}$, to oznaczamy ten fakt $\mathcal{M} \models \varphi$ (mówimy też, że $\mathcal{M}$ *spełnia* φ). W przeciwnym przypadku formuła ta jest *fałszywa*, co oznacza, że $\mathcal{M}$ spełnia jej negację: $\mathcal{M} \models \neg\varphi$ (symbol $\neg$ to symbol negacji logicznej). I tak w każdym modelu, w którym umiemy dodawać i mnożyć, albo istnieje liczba, której kwadrat jest równy 1+1, albo taka liczba nie istnieje. Czyli zawsze albo $\mathcal{M} \models \varphi$, albo $\mathcal{M} \models \neg\varphi$. Powyższą notację możemy rozszerzyć na zbiory formuł: $\mathcal{M} \models \Gamma$, jeśli $\mathcal{M} \models \varphi$ dla każdej formuły φ z Γ. Relację $\models$ nazywamy relacją *semantyczną*, bo opisuje ona faktyczne światy matematyczne – modele.

Jedynkę można zdefiniować jako element neutralny mnożenia.

Opisany powyżej sposób pracy matematyka dzieli się zatem na dwa etapy: w pierwszym formułujemy aksjomatykę Γ i przekonujemy się, że $\mathcal{M} \models \Gamma$; w drugim próbujemy wywnioskować interesującą nas formułę φ z Γ. Wnioskowanie takie to podanie ścisłego matematycznego dowodu: pewnego skończonego obiektu matematycznego, który na podstawie przyjętych reguł wnioskowania i założenia z Γ wykazuje φ. Gdy taki dowód istnieje, to piszemy $\Gamma \vdash \varphi$. Zauważmy, że ta relacja $\vdash$ nie odwołuje się do żadnego modelu $\mathcal{M}$ – nazywamy ją *konsekwencją syntaktyczną*, gdyż zarówno formuły z Γ, formuła φ, jak i sam dowód to w ostateczności skończone napisy.

Matematyka nie jest jednak przecież tylko systemem formalnym, sprowadzającym się do przeszukiwania wszystkich możliwych dowodów. Pracując nad interesującym nas twierdzeniem, opieramy się na intuicjach dotyczących faktycznych modeli. Oznacza to, że tak naprawdę interesuje nas relacja *konsekwencji semantycznej*: powiemy, że φ jest *konsekwencją semantyczną* aksjomatów Γ (ozn. $\Gamma \models \varphi$), jeśli każdy model $\mathcal{M}$, który spełnia wszystkie aksjomaty z Γ, spełnia też φ.

Jak zauważyliśmy powyżej, każdy konkretny model $\mathcal{M}$ musi albo spełniać daną formułę ($\mathcal{M} \models \varphi$), albo jej nie spełniać, co oznacza, że $\mathcal{M} \models \neg\varphi$. Własność ta przestaje być prawdą dla powyższej relacji semantycznej konsekwencji: może być tak, że pewna aksjomatyka Γ ma dwa istotnie różne modele $\mathcal{M}_1$ i $\mathcal{M}_2$, takie że $\mathcal{M}_1 \models \Gamma$ oraz $\mathcal{M}_2 \models \Gamma$. Może wtedy istnieć pewna formuła φ, która zachodzi w modelu $\mathcal{M}_1$, natomiast jest fałszywa w modelu $\mathcal{M}_2$. Wtedy zarówno $\Gamma \not\models \varphi$, ale też $\Gamma \not\models \neg\varphi$, co oznacza, że formuła φ jest *niezależna* od Γ. Przykładem takiej sytuacji jest przypadek, gdy Γ to standardowy zbiór pięciu aksjomatów Euklidesa, zaś φ to tzw. aksjomat Pascha, sformułowany przez Moritza Pascha w 1882 roku. Fakt, że sytuacja taka ma czasami miejsce, nie jest może szokujący, bo nietrudno wyobrazić sobie, że ktoś zapomni o jakimś dość istotnym aksjomacie, powodując, że teoria Γ pewnych własności modeli po prostu nie określa. Jak się później przekonamy, sytuacja jest dużo bardziej poważna, i nie chodzi tu tylko o nieuwagę przy doborze Γ.

Naszym celem jest teraz zrozumienie zależności pomiędzy dwiema relacjami konsekwencji: semantycznej $\models$ oraz syntaktycznej $\vdash$. Po pierwsze, reguły, jakich używa się w konstrukcji dowodów matematycznych, są same w sobie poprawne, co daje nam następujące twierdzenie o poprawności.

Twierdzenie 1 (o poprawności). *Jeśli* $\Gamma \vdash \varphi$, *to* $\Gamma \models \varphi$.

Po drugie okazuje się, że zachodzi też twierdzenie odwrotne, wykazane przez Kurta Gödela w 1929 roku.

Twierdzenie 2 (o pełności). *Jeśli* $\Gamma \models \varphi$, *to* $\Gamma \vdash \varphi$.

Mówi ono, że jeśli pewna własność φ jest prawdziwa we wszystkich modelach danej aksjomatyki, to musi mieć pewien skończony dowód.

Powyższe twierdzenie daje nam dużą dozę optymizmu: o ile tylko teza φ, którą chcemy udowodnić, faktycznie wynika z przyjętych założeń Γ, to musi o tym zaświadczać pewien skończony dowód, który prędzej czy później znajdziemy. Optymizm ten jednak jest zwodniczy, o czym może świadczyć historia Georga Cantora. Jednym z głównych obiektów jego badań była sformułowana przez niego w 1878 roku Hipoteza Continuum (ozn. φ_{CH}). Hipoteza ta mówi, że każdy podzbiór X liczb rzeczywistych jest albo przeliczalny (jego elementy daje się ponumerować liczbami naturalnymi), albo równoliczny ze wszystkimi liczbami rzeczywistymi (istnieje bijekcja pomiędzy X a $\mathbb{R}$). Cantor włożył ogromny wysiłek w próby wykazania lub obalenia tej hipotezy. Niestety nie udało mu się osiągnąć żadnego z tych celów, co podobno doprowadziło go do szaleństwa.

Z obecnej perspektywy badania Cantora można interpretować jako próby sprawdzenia, czy formuła φ_{CH} jest konsekwencją semantyczną przyjętej aksjomatyki teorii mnogości ZFC, czyli aksjomatyki Zermelo–Fraenkla wraz z aksjomatem wyboru (ozn. Γ_{ZFC}). Jak się okazało, porażka Cantora nie wynikała z braku jego pomysłowości. Najpierw, w 1940 roku, Kurt Gödel stworzył model $\mathcal{M}_1$ spełniający aksjomaty Γ_{ZFC}, w którym φ_{CH} zachodzi. Oznacza to, że $\Gamma_{\mathrm{ZFC}} \nvdash \neg\varphi_{\mathrm{CH}}$, co tłumaczy, dlaczego Cantorowi nie udało się obalić Hipotezy Continuum. Problem ten zamknął Paul Cohen w 1963 roku, konstruując przy użyciu stworzonej przez siebie metody forcingu inny model, $\mathcal{M}_2$, spełniający aksjomaty Γ_{ZFC}. W modelu tym Hipoteza Continuum jest fałszywa, co pokazuje,

Rozwiązanie zadania M 1670.

$$
\begin{aligned}
1 &= [\sqrt{\pi}],\\
2 &= -[-\sqrt{\pi}],\\
3 &= [\pi],\\
4 &= -[-\pi],\\
5 &= \left[\sqrt{\sqrt{([\pi]!)!}}\,\right],\\
6 &= [\pi]!,\\
7 &= -[-\sqrt{([\pi]!)!!}\,],\\
8 &= -([-\pi]+[-\pi]),\\
9 &= [\pi\cdot\pi],\\
10 &= \left[\sqrt{[\sqrt{\sqrt{([\pi]!)!}}\,]!}\,\right],\\
11 &= -\left[-\sqrt{[\sqrt{\sqrt{([\pi]!)!}}\,]!}\,\right],\\
12 &= \Big[-\pi\cdot[-\pi]\Big],\\
13 &= -\Big[\pi\cdot[-\pi]\Big],\\
14 &= [-(\pi+\pi)]\cdot[-\sqrt{\pi}],\\
15 &= [\pi\cdot\sqrt{\pi}]\cdot[\pi],\\
16 &= [-\pi]\cdot[-\pi],\\
17 &= [\pi\cdot\pi\cdot\sqrt{\pi}],\\
18 &= [\pi+\pi]\cdot[\pi],\\
19 &= [(\pi+\pi)\cdot\pi],\\
20 &= -[-(\pi+\pi)\cdot\pi].
\end{aligned}
$$

Dowód twierdzenia 1: Załóżmy, że istnieje pewien skończony dowód P formuły φ oparty na aksjomatach z Γ. Weźmy dowolny model $\mathcal{M}$ i załóżmy, że $\mathcal{M} \models \Gamma$. Przez indukcję po strukturze dowodu P wykazujemy, że wszystkie pośrednie występujące w nim formuły również są prawdziwe w $\mathcal{M}$. W takim razie $\mathcal{M} \models \varphi$.

Schemat konstrukcji z twierdzenia 2: Rozumujemy przez sprzeczność, zakładając, że $\Gamma \nvdash \varphi$. Najpierw rozszerzamy naszą aksjomatykę do większego zbioru $\Gamma' \supseteq \Gamma$. Robimy to, dorzucając do Γ kolejne formuły, dbając, by na każdym etapie wciąż zachodziło $\Gamma' \nvdash \varphi$. Następnie tworzymy tzw. *model syntaktyczny* $\mathcal{M}_0$, który ma tę własność, że dla każdej formuły ψ zachodzi $\mathcal{M}_0 \models \psi$ wtedy i tylko wtedy, gdy $\Gamma' \vdash \psi$. Oznacza to w szczególności, że $\mathcal{M}_0 \models \Gamma' \supseteq \Gamma$ oraz $\mathcal{M}_0 \not\models \varphi$. Czyli $\Gamma \not\models \varphi$ i mamy poszukiwaną sprzeczność.

Aksjomat wyboru mówi, że w każdej rodzinie zbiorów niepustych można wybrać po jednym elemencie z każdego zbioru. Pomimo swego niegroźnego sformułowania, jego założenie implikuje istnienie różnych „dziwnych” obiektów, jak np. zbiorów niemierzalnych.

Rozwiązania zadań z artykułu *O pewnej metodzie rozwiązywania równań „nierozwiązywalnych”.*

Zadanie 1.

$$xe^{\frac{x}{2}} = \sqrt{2},$$
$$\frac{x}{2}e^{\frac{x}{2}} = \frac{1}{\sqrt{2}},$$
$$x = 2W(1/\sqrt{2}).$$

Zadanie 2. Niech $y = 2 - x$. Wtedy $ye^y = e^2$ i stąd $y = W(e^2)$, czyli $x = 2 - W(e^2)$.

Podobnie rozwiązujemy **zadanie 3**, które jest ogólniejszą wersją zadania 2.

Zadanie 4. Postać równoważna równania

$$\ln x = a + \frac{b}{x}$$

to

$$\ln \frac{x}{e^a} = \frac{b}{e^a} : \frac{x}{e^a}.$$

Dalej dokonujemy przekształceń:

$$\ln \frac{x}{e^a} e^{\ln \frac{x}{e^a}} = \frac{b}{e^a},$$
$$\ln \frac{x}{e^a} = W(b/e^a),$$
$$x = e^{a+W(b/e^a)}.$$

że $\Gamma_{\mathrm{ZFC}} \not\vdash \varphi_{\mathrm{CH}}$, czyli że aksjomaty ZFC nie są dość silne, by wykazać prawdziwość Hipotezy Continuum. Czyli hipoteza ta jest od nich niezależna!

Naturalną reakcją na przedstawiony powyżej obraz sytuacji jest stwierdzenie, że podobnie jak pięć aksjomatów Euklidesa było „za słabych”, pozwalając, by aksjomat Pascha był od nich niezależny, może również aksjomaty ZFC są po prostu „zbyt słabe”. Prowadzi to do pojęcia teorii *zupełnej*, czyli takiego zbioru formuł Γ, że dla każdej formuły φ zachodzi albo $\Gamma \models \varphi$, albo $\Gamma \models \neg\varphi$. Przykładem takiej teorii może być Γ' użyta w dowodzie twierdzenia 2, mogą być aksjomaty niepustego gęstego porządku liniowego bez elementu minimalnego ani maksymalnego (opisujące $\langle \mathbb{Q}, \leqslant \rangle$) czy też aksjomaty ciała rzeczywiście domkniętego (opisujące $\langle \mathbb{R}, (+), (\cdot) \rangle$). Warto może dodać, że ostatni z wymienionych wyników był uzyskany przez Alfreda Tarskiego w 1931 roku.

Chcąc uniknąć problemów, na jakie natrafił Cantor, można by postulować, aby zamiast teorii ZFC przyjąć za podstawy matematyki jakąś silniejszą teorię, o której wiedzielibyśmy, że jest zupełna. Okazuje się, że wtedy moglibyśmy wręcz zautomatyzować proces sprawdzania prawdziwości formuł. Przecież skoro $\Gamma \models \varphi$ albo $\Gamma \models \neg\varphi$, to na mocy twierdzenia o pełności również $\Gamma \vdash \varphi$ albo $\Gamma \vdash \neg\varphi$. Wystarczy więc generować kolejne napisy przypominające kształtem dowody i sprawdzać, czy przypadkiem dany napis nie jest dowodem, że φ lub że $\neg\varphi$. Wówczas po skończenie wielu krokach musimy natknąć się na dowód jednego z tych dwóch faktów i wtedy wiemy, czy φ zachodzi!

Podstawowym ryzykiem wzmacniania rozważanych teorii jest to, że mogą stać się wewnętrznie sprzeczne: teoria Γ jest *sprzeczna*, jeśli daje się w niej udowodnić fałsz (ozn. $\perp$). Twierdzenia o poprawności i pełności pozwalają scharakteryzować teorie sprzeczne: $\Gamma \vdash \perp$ wtedy i tylko wtedy, gdy Γ nie ma żadnego modelu (bo w każdym modelu $\mathcal{M}$ zachodzi $\mathcal{M} \not\models \perp$). Oczywiście nie chcielibyśmy pracować w takiej teorii, gdyż nie opisuje ona żadnego modelu. W historii matematyki zdarzało się już, że proponowano sprzeczne rozszerzenia pewnych teorii: na przykład istnienie liczby kardynalnej Reinhardta (zaproponowane przez Williama Reinhardta w 1967 roku) okazało się sprzeczne z teorią ZFC; wykazał to Herbert Kunen w 1971 roku. Pokazuje to, że potrzebna jest spora ostrożność w szukaniu takich wzmocnień.

Niestety sytuacja nie jest tak prosta, jak moglibyśmy liczyć, i niezależnie od naszej ostrożności nie mamy szans na znalezienie zupełnej teorii stanowiącej podstawy matematyki. Wyjaśnia to drugie twierdzenie Kurta Gödla z 1931 roku.

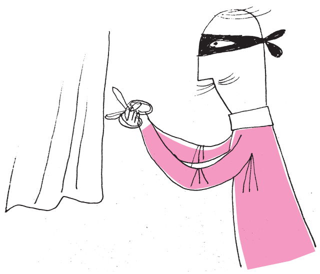
Twierdzenie 3 (o niezupełności). *Jeśli teoria Γ jest dostatecznie silna, by zdefiniować w niej arytmetykę $\langle \mathbb{N}, (+), (\cdot) \rangle$, wszystkie aksjomaty Γ daje się konstruktywnie wyliczać* oraz Γ jest niesprzeczna, to istnieje formuła ψ niezależna od Γ (czyli $\Gamma \not\models \psi$ oraz $\Gamma \not\models \neg\psi$).*

*Nawet jeśli tych aksjomatów jest nieskończenie wiele! Słowo „konstruktywnie” należy tu rozumieć przez opis jakąś wspólną formułą lub – bardziej ściśle – przez możliwość ich enumerowania przez program komputerowy.

Dowód tego twierdzenia bazuje mocno na teorii obliczalności i możliwości zakodowania w rozważanej teorii paradoksu kłamcy – paradoksu, w którym pewien człowiek wypowiada zdanie: *mówiąc to zdanie, kłamię*. I, podobnie jak w przypadku tego paradoksu, tak skonstruowane zdanie nie może być ani prawdą, ani kłamstwem (fałszem). By powyższe kodowanie było wykonalne, konieczna jest możliwość mówienia w obrębie samej teorii o tym, czy potrafi ona czegoś dowieść. W szczególności można napisać formułę $\varphi_{\mathrm{CON}(\Gamma)}$ wyrażającą fakt, że Γ jest niesprzeczna (czyli $\Gamma \not\vdash \perp$) – formuła ta mówi, że nie istnieje obiekt arytmetyczny, który koduje w sobie skończony dowód fałszu z aksjomatów w Γ. Jak się okazuje, za formułę ψ ze sformułowania twierdzenia 3 można przyjąć właśnie $\varphi_{\mathrm{CON}(\Gamma)}$ – **oznacza to, że żadna dostatecznie silna teoria nie jest w stanie udowodnić ani obalić swojej własnej niesprzeczności!**

Warto może zauważyć, że twierdzenie 3 nie stoi w sprzeczności z faktem, że teoria ciał rzeczywiście domkniętych (opisująca model liczb rzeczywistych z dodawaniem i mnożeniem $\langle \mathbb{R}, (+), (\cdot) \rangle$) jest zupełna – teoria ta jest zbyt słaba, by wyróżnić spośród wszystkich liczb rzeczywistych liczby naturalne $\mathbb{N}$, więc nie spełnia pierwszego z założeń na temat Γ.

Na zakończenie warto zwrócić uwagę na bogactwo słownictwa w języku polskim: dzięki rozróżnieniu słów *pełność* i *zupełność* dwa twierdzenia Gödla nazywają



się twierdzeniami o *pełności* i *niezupełności*. Sytuacja ma się inaczej w języku angielskim, gdzie twierdzenia te nazywają się *Gödel's completeness theorem* oraz *Gödel's incompleteness theorem*, sugerując, że Gödel zwariował i raz udowodnił *completeness*, a kolejnym razem *incompleteness*.

Podsumowując, o ile w konkretnym modelu zachodzi albo $\mathcal{M} \models \varphi$, albo $\mathcal{M} \models \neg\varphi$, to w przypadku teorii może się zdarzyć, że $\Gamma \not\models \varphi$ oraz $\Gamma \not\models \neg\varphi$. Co gorsze, wszystkie dostatecznie silne teorie mają takie zdania niezależne – pochodzą one od pomysłowego kodowania w danej teorii paradoksu kłamcy i wykorzystania narzędzi z teorii obliczeń. Oznacza to, że zawsze musimy się liczyć z tym, iż rozważana formuła może leżeć na tej „ziemi niczyjej": nie da się jej ani udowodnić, ani obalić. Z drugiej strony, nasza sytuacja jest tak dobra, jak tylko w tych warunkach może być: jeśli tylko dana formuła φ jest konsekwencją Γ we wszystkich modelach, to prędzej czy później znajdziemy na to dowód.

Przez firankę, czyli jak odkryto pulsary

*Aleksander SCHWARZENBERG-CZERNY**

* Centrum Astronomiczne im. Mikołaja Kopernika w Warszawie

Gdy zapada zmrok i za oknem zapalają się światła, warto spojrzeć na odległą latarnię przez gęstą firankę lub półprzezroczystą zasłonę, rozpostartą oburącz prostopadle do kierunku widzenia. Obraz latarni rozpadnie się na konstelację światełek rozłożonych w kratkę, zorientowaną równolegle do nitek materiału. A teraz przysuwając jedną z rąk nieco bliżej twarzy ustawmy materiał skosem do kierunku patrzenia: światełka zaczynają się rozsuwać. Zaraz, zaraz, przecież (pionowe) nitki widziane pod kątem są ułożone bliżej, więc skąd rozsunięcie?

Być może to właśnie z powodu braku jasnych świateł Izaak Newton pisząc swe dzieło o optyce, opisał światło jako strumień cząsteczek. Interferencję można bowiem wytłumaczyć tylko opierając się na falowej naturze światła.

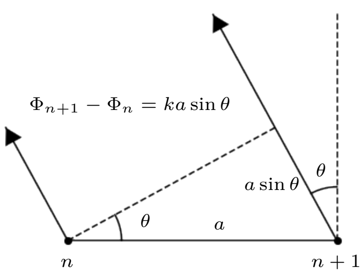
Dyfrakcja światła ze źródeł punktowych. Sinusoidalną falę światła, a dokładniej wektor pola elektrycznego E o amplitudzie E_0, wygodnie jest przedstawić za de Moivre'em jako $E = E_0 e^{i\Phi} \equiv E_0(\cos\Phi + i\sin\Phi)$, gdzie faza fali Φ zależna od czasu t i położenia l w kierunku rozchodzenia się fali to $\Phi(t,l) = 2\pi(t/T - l/\lambda) \equiv \omega t - kl$. Stałe T i λ to okres i długość fali, $\omega = 2\pi/T$ i $k = 2\pi/\lambda$ to częstość kołowa i liczba falowa, a prędkość fali $c = \lambda/T = \omega/k$. Przez i oznaczamy urojoną jedność $i^2 = -1$ i tylko część rzeczywistą rozwiązania (niezawierającą i) traktujemy jako prawdziwą falę. Rozpatrzmy teraz N identycznych źródeł fali rozłożonych wzdłuż linii prostej w odstępach a. Jeśli oscylują one w fazie, ale patrzymy na nie z oddali pod kątem θ względem prostopadłej, to różnica drogi do obserwatora pomiędzy sąsiednimi źródłami będzie $l = a\sin\theta$, a odpowiednia różnica fazy to

O użyteczności wzoru de Moivre'a świadczy m.in. poniższa regułka, pozwalająca zapamiętać wzory trygonometryczne dla sumy kątów: $\cos(\alpha+\beta) + i\sin(\alpha+\beta) = e^{i\alpha+i\beta} =$ $= (\cos\alpha + i\sin\alpha)(\cos\beta + i\sin\beta) =$ $= \cos\alpha\cos\beta - \sin\alpha\sin\beta +$ $+ i(\cos\alpha\sin\beta + \sin\alpha\cos\beta)$. Porównanie osobno części rzeczywistej i urojonej daje wspomniane wzory.

(1) $$\Phi_{n+1} - \Phi_n \equiv \phi = ka\sin\theta.$$

Zatem obserwowana suma pól elektrycznych to

(2) $$E = E_0 e^{i\Phi_1} \sum_{n=0}^{N-1} e^{in\phi} = E_0 e^{i\Phi_1} \frac{1-e^{iN\phi}}{1-e^{i\phi}} = E_0 e^{i(\Phi_1 + N\phi/2 - \phi/2)} \frac{\sin(N\phi/2)}{\sin(\phi/2)},$$



gdzie Φ_1 to obserwowana faza pierwszego źródła. Obliczając sumę, skorzystaliśmy ze wzoru na sumę ciągu geometrycznego. Łatwo go również wykazać indukcyjnie, bowiem po dodaniu kolejnego wyrazu i sprowadzeniu do wspólnego mianownika dostaniemy wyrażenie dla $N+1$ źródeł. Przyjmując za jednostkę natężenie pola dla $\theta = \phi = 0$, otrzymujemy

(3) $$I_1(\theta) = \frac{\sin^2(N\phi/2)}{N^2\sin^2(\phi/2)} \xrightarrow[\substack{N\to\infty \\ Na \equiv D = \text{const}}]{} \frac{\sin^2(\Delta\Phi/2)}{(\Delta\Phi/2)^2} \equiv I_2(\theta).$$

Z wzoru de Moivre'a wynika też $e^{i\alpha} - e^{-i\alpha} = 2i\sin\alpha$. Obserwator nie rejestruje jednak pola elektrycznego E, a natężenie promieniowania proporcjonalne do uśrednionego po czasie kwadratu rzeczywistej części pola $I \sim \langle(\operatorname{Re} E)^2\rangle$. Skoro po czasie $T/4$ rzeczywista i urojona część Φ zamieniają się rolami, to $\langle(\operatorname{Re} E)^2\rangle = \langle(\operatorname{Im} E)^2\rangle = \frac{1}{2}\langle|E|^2\rangle = \text{const}$, gdzie ostatnia równość wynika z faktu, że moduł jedynego czynnika zmiennego w czasie to $|e^{i\omega t}|^2 = 1 = \text{const}$ z jedynki trygonometrycznej.

Funkcja I_1 opisuje natężenie światła w obrazie dyfrakcyjnym od N spójnych (zgodnych w fazie) źródeł, na przykład za układem wąskich szczelin oświetlonych światłem lasera lub odległej lampy. Dla $\phi \approx 0$ sinusy można zastąpić ich argumentami i widać, że natężenie osiąga maksimum główne $I_1(0) = 1$. Graniczny rozkład I_2 po prawej stronie odpowiada dążącej do nieskończoności liczby źródeł $N \to \infty$ umieszczonych w przedziale o ustalonej szerokości $\text{const} = D \equiv Na$, czyli przy a i θ dążących do 0. Sinus w mianowniku można wtedy zastąpić jego argumentem: oznaczając różnicę skrajnych faz $\Delta\Phi \equiv N\phi = Nka\sin\theta = kD\sin\theta$,