Zasada Fermata, czyli o przyczynowości i celowości w fizyce

Dr hab. Andrzej SZYMACHA

Zapewne większość Czytelników zna, przynajmniej powierzchownie, zasadę Fermata. Mówi ona, że światło spośród wszystkich możliwych torów łączących dwa ustalone punkty wybiera ten, którego przebycie wymaga najkrótszego czasu. Formalne znaczenie tej zasady, bez wnikania w głębsze subtelności, polega na tym, że ujmuje ona bardzo elegancko wszystkie prawa optyki geometrycznej. Wykazanie tego jest dość proste. Przypomnijmy tylko, że dla wszystkich zastosowań optyki geometrycznej (np. konstrukcja obrazów w różnych przyrządach) wystarczy wiedzieć, że

a) w ośrodku jednorodnym światło rozchodzi się po liniach prostych,

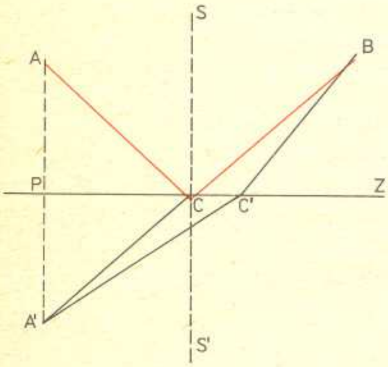
b) w przypadku odbicia światła kąt padania jest równy kątowi odbicia,

c) przekraczając powierzchnię oddzielającą dwa ośrodki, w których światło rozchodzi się z różnymi prędkościami promień światła załamuje się zgodnie z prawem Snelliusa:

$$\frac{\sin\alpha}{\sin\beta}=\frac{n_2}{n_1}, \qquad \text{gdzie } n_i=\frac{c}{v_i}=\frac{\text{prędkość światła w próżni}}{\text{prędkość światła w danym ośrodku}}.$$

Wykażemy teraz, że z zasady Fermata wynikają prawa **a**, **b** i **c**.

a) W ośrodku jednorodnym współczynnik załamania n, a zatem i prędkość światła v są stałe i czas przebiegu jest wprost proporcjonalny do odległości. Zasada najkrótszego czasu jest zatem równoważna w tym przypadku zasadzie najmniejszej odległości. Jest oczywistym faktem geometrycznym, że właśnie odcinek prostej łączący dane punkty A i B charakteryzuje się najmniejszą długością w porównaniu z innymi liniami łączącymi te punkty.



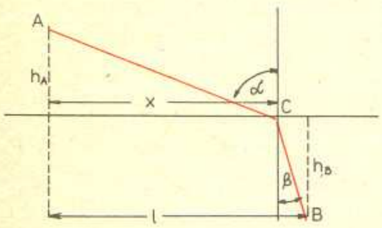
b) Odbicie. Rozpatrzmy długość toru promienia od A do B wzdłuż drogi, która choć w jednym punkcie dotyka powierzchni zwierciadła. Zgodnie z poprzednimi wynikami wiemy, że od punktu A do zwierciadła i od zwierciadła do punktu B światło poruszać się będzie po prostej. Wyznaczenie toru promienia sprowadza się w tym wypadku do wyznaczenia punktu C', dla którego łamana $AC'B$ będzie najkrótsza. W tym celu zauważmy, że długość łamanej $AC'B$ jest identyczna z długością łamanej $A'C'B$, jeśli punkt A' jest względem zwierciadła położony symetrycznie do punktu A. Łamana $A'C'B$ jest najkrótsza wtedy, gdy jest odcinkiem $A'B$. Jeżeli $A'CB$ jest odcinkiem, to

$$\sphericalangle A'CS' = \sphericalangle SCB. \tag{1}$$

Ponieważ ΔAPC jest przystający do $\Delta A'PC$, to $\sphericalangle PCA = \sphericalangle PCA'$.
Kąty dopełniające tych kątów są też równe.

$$\sphericalangle A'CS' = \sphericalangle ACS. \tag{2}$$

Porównując równości (1) i (2) widzimy, że $\sphericalangle SCB = \sphericalangle ACS$.
A to właśnie chcieliśmy wykazać.



c) Załamanie. I w tym wypadku najkrótszy czas przebiegu realizuje się na pewno dla toru będącego łamaną (po każdej stronie płaszczyzny załamującej ośrodek jest z założenia jednorodny). W celu wyznaczenia warunku na położenie punktu C, przy którym czas jest najkrótszy, obliczmy ten czas dla położenia C scharakteryzowanego odległością x, a następnie zminimalizujmy to wyrażenie ze względu na x.

$$T=\frac{AC}{v_1}+\frac{CB}{v_2}=\frac{\sqrt{h_A^2+x^2}}{\frac{c}{n_1}}+\frac{\sqrt{h_B^2+(l-x)^2}}{\frac{c}{n_2}}.$$

W celu wyznaczenia minimum przyrównajmy do zera pochodną $\frac{\mathrm{d}T}{\mathrm{d}x}$.

$$0=\frac{\mathrm{d}T}{\mathrm{d}x}=\frac{1}{c}\left[n_1\frac{\mathrm{d}}{\mathrm{d}x}\sqrt{h_A^2+x^2}+n_2\frac{\mathrm{d}}{\mathrm{d}x}\sqrt{h_B^2+(l-x)^2}\right]=$$

$$=\frac{1}{c}\left[\frac{n_1x}{\sqrt{h_A^2+x^2}}-\frac{n_2(l-x)}{\sqrt{(l-x)^2+h_B^2}}\right]=\frac{1}{c}\left[n_1\frac{x}{AC}-n_2\frac{l-x}{CB}\right].$$

Ponadto

$$\frac{x}{AC}=\sin\alpha,\quad \frac{l-x}{CB}=\sin\beta. \text{ Zatem } n_1\sin\alpha=n_2\sin\beta \text{ i } \frac{\sin\alpha}{\sin\beta}=\frac{n_2}{n_1}=n_{12}.$$

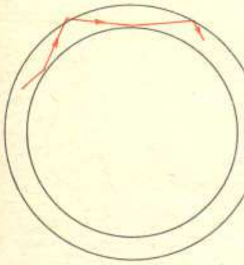
Widzimy więc, że istotnie zasada najkrótszego czasu prowadzi do prawa Snelliusa.

Odkrycie zasady Fermata wywołało w swoim czasie poważny spór filozoficzny: Czy Przyroda działa przyczynowo, czy celowo? Gdybyśmy znali tylko zasadę Fermata, bez wniosków z niej wynikających, to skłanialibyśmy się do drugiego poglądu. Przyroda wybiera dla promienia taki przebieg, by wypełnić pewien cel globalny, zużyć na przejście światła jak najmniej czasu. Gdyby od dziecka wychowywano nas w tym przekonaniu, to łatwo uznalibyśmy je w końcu za oczywiste i nie wymagające dalszych komentarzy. Z drugiej strony prawa optyki **a**, **b**, **c** mają charakter lokalny, przyczynowy. Prawo **a** mówi, że światło wysyłane z punktu A

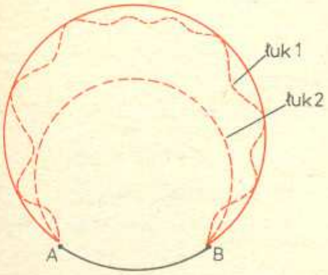
w określonym kierunku porusza się nieco do przodu i nie napotykając żadnej przeszkody kontynuuje swój prostoliniowy bieg na podobieństwo swobodnego punktu materialnego. Powierzchnie odbijające bądź załamujące zmieniają kierunek tego biegu na bardzo małym obszarze (lokalnie) według określonych praw, prowadząc krok po kroku do takiego, a nie innego ostatecznego biegu promienia.

Dylemat powyższy nazwałem filozoficznym raczej niż fizycznym, gdyż, jak wskazują zanalizowane przykłady, oba sformułowania są matematycznie równoważne. Jednakże, gdy przejdziemy do sytuacji bardziej skomplikowanych, jedno podejście może okazać się bardziej wartościowe od drugiego. We współczesnej fizyce uważa się, że rozumowanie przyczynowe, lokalne, wnika istotnie w naturę rzeczy, podczas gdy efekt globalny jest tylko uboczną konsekwencją, którą zresztą warto lepiej zrozumieć.

Argumentów za tym, że podejście przyczynowe jest pierwotne, a efekt globalny (w postaci najkrótszego czasu) wtórny, można dostarczyć pozostając w ramach optyki geometrycznej, ale stają się one szczególnie dobitne jeśli uwzględnić, że optyka geometryczna jest tylko pewnym przypadkiem granicznym (idealizacją) optyki falowej.

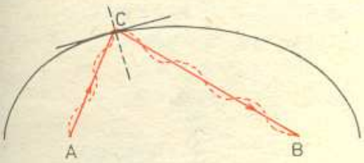


W artykule tym ograniczymy się do argumentów pierwszego rodzaju. Polegają one na zauważeniu, że czas przebiegu światła — wbrew podanemu powyżej sformułowaniu — nie zawsze jest najkrótszy. Rozpatrzmy następujący przykład. Mamy dwie odbijające koncentryczne powierzchnie sferyczne i rozpatrujemy ruch światła między nimi. Przechodząc z odległością między sferami do zera uzyskamy graniczny przypadek promienia światła zmuszonego poruszać się (ze stałą prędkością) po powierzchni kuli. Nie ulega wątpliwości, że w tym granicznym przypadku kształty promieni będą pokrywały się z łukami kół wielkich na powierzchni kuli (Czytelnik powinien postarać się wykazać to formalnie). Jest to zgodne z zasadą Fermata, ale tylko dla łuków niezbyt dużych. Dla łuków większych od połowy obwodu koła wielkiego łuk AB wcale nie jest najkrótszą odległością między punktami A i B, ale na pewno jest to możliwy tor promienia. W otoczeniu tego łuku istnieją zarówno łuki dłuższe (1) jak też krótsze (2). Widać, że z celowością jesteśmy w tym przypadku na bakier. Prawdą natomiast jest, że po podzieleniu tego wielkiego łuku na mniejsze fragmenty uzyskujemy dla każdego z tych fragmentów prawdziwe minimum. Dowodzi to, że prawa lokalne są zawsze spełnione — natomiast minimum czasu przebiegu jest ich matematyczną konsekwencją jedynie dla torów dostatecznie krótkich.



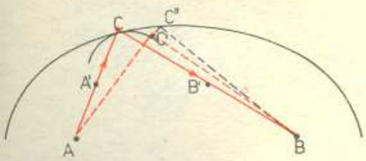
A oto inny, bardziej realistyczny przykład.

Rozpatrzmy elipsoidę obrotową. Matematyczną własnością elipsy jest, że $r_1 + r_2 = \text{const}$ (jeśli A i B są ogniskami elipsy). Prawdą jest również, że styczna do elipsy jest prostopadła do dwusiecznej kąta ACB. Zatem łamana ACB jest jednym z możliwych torów promienia światła. Zgodnie z pierwszą własnością w otoczeniu tego toru istnieją tory o identycznej długości (a zatem o identycznym czasie przebiegu). Istnieją również tory o dłuższym czasie przebiegu (linie krzywe, np. ta zaznaczona linią przerywaną). Znów zasada minimum nie działa.



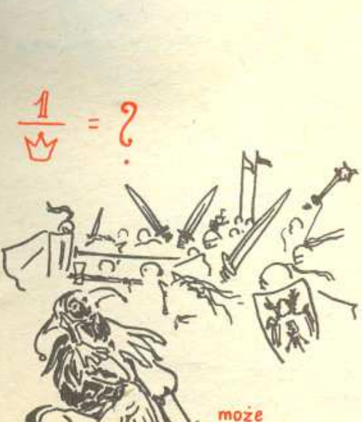
Jest ona naruszona jeszcze wyraźniej, jeśli wprowadzić zwierciadło kuliste Z styczne do elipsoidy w punkcie C. Linia ACB jest znów możliwym torem światła (spełnia prawo odbicia). Ale w otoczeniu łamanej ACB istnieją tory krótsze od niej. Takim torem jest łamana $AC'B$. Z rysunku widać, że

$$AC'B < AC''B = ACB.$$



I znów, podobnie jak w przypadku z powierzchnią kuli, można dowieść, że dla punktów A' i B' dostatecznie bliskich łamana $A'CB'$ jest najkrótsza. A więc dla małych fragmentów toru zasada Fermata jest słuszna, a globalnie nie. Podważa to pogląd o jakiejkolwiek „celowości" i świadczy na rzecz przyczynowości.

Pozostaje do wyjaśnienia, dlaczego w rozpatrywanych na początku przykładach (**a, b, c**) zasada Fermata okazała się równoważna prawom optyki geometrycznej. Otóż we wszystkich tych trzech przykładach nie było w geometrii problemu niczego, co definiowałoby jakąś skalę odległości. Tym samym odcinki o dowolnych długościach są równoprawne, każdy może być zatem uważany za „dostatecznie krótki". Ograniczając się do płaskich powierzchni odbijających i załamujących udowodniliśmy więc w istocie słuszność zasady Fermata dla lokalnych fragmentów toru. Kiedy przeszliśmy do kul i elipsoid, owa równoważność się załamała.



Na zakończenie dodam, że choć zasada Fermata w najprostszej oryginalnej wersji okazała się za mało ogólna (i w sensie globalnym nieprawdziwa), to istnieje jej uogólnienie wolne od owych ograniczeń. W tym ogólniejszym sformułowaniu zasady Fermata należy słowo „najmniejszy" zastąpić słowem „stacjonarny". Mówimy, że czas przebiegu dla jakiegoś toru jest stacjonarny, jeśli dla bliskich mu torów czas przebiegu różni się mało. Tor minimalny jest torem stacjonarnym, ale nie na odwrót. Owa stacjonarność odgrywa decydującą rolę przy analizowaniu zasady Fermata z punktu widzenia optyki falowej. Tor stacjonarny ma tę własność, że w jego otoczeniu istnieją tory, wzdłuż których propagujące się (zgodnie z zasadą Huyghensa) fale interferują konstruktywnie (mała zmiana czasu przelotu odpowiada małej zmianie fazy fali). Otoczenie torów stacjonarnych optyki geometrycznej odgrywa więc szczególną i wyróżnioną rolę w obrazie falowym światła. Zasada Fermata (uogólniona) jest tego naturalną konsekwencją. Ale podejście falowe jest lokalne, przyczynowe, a nie celowe.