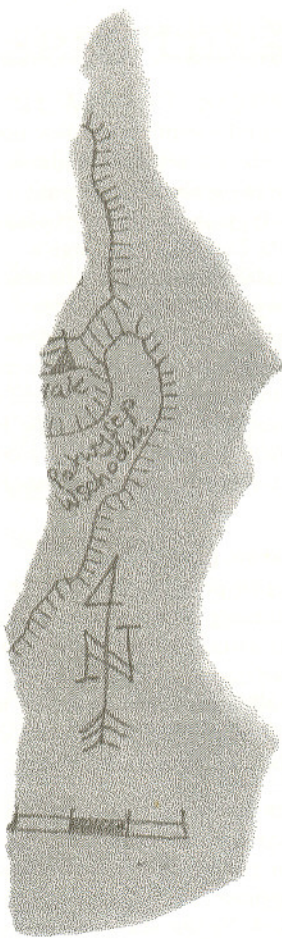


Zasada najmniejszego działania w fizyce

Krzysztof MEISSNER

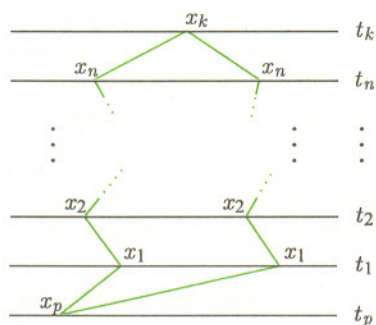
Gdy gramy w koszykówkę, zdarza się (niestety częściej niż rzadziej), że rzucając piłkę nie trafimy do kosza – dla nikogo nie ulega wtedy wątpliwości, że źle rzuciliśmy (pod złym kątem albo z za małą prędkością) i nie możemy się tłumaczyć, że mimo dobrego rzutu trajektoria „złośliwie” odchyliła się od prawidłowej. Wynika to z powszechnego przekonania, że ruch piłki jest jednoznacznie określony (pomijając opór powietrza i tym podobne małe poprawki) przez miejsce rzutu i nadaną jej prędkość początkową. Przyjęcie jako ogólnie prawdziwego takiego niezmiennego postulatu, określającego jednoznacznie ruch, niesie ze sobą niezwykle głębokie konsekwencje zarówno praktyczne, jak i filozoficzne. Szczególnie w wieku XIX wyciągano z istnienia takiego postulatu daleko idące wnioski co do determinizmu w świecie, jednak mechanika kwantowa dowodzi, że w świetle zasady Heisenberga ruchu jednoznacznie przewidzieć się nie da i wnioski te opierały się na błędnych założeniach. Jednak na szczęście w życiu codziennym taki postulat jest wystarczająco dobrym przybliżeniem. Każdy może sobie wyobrazić przykłady, co by się działo, gdyby ten postulat nie obowiązywał i np. gorąca woda z czajnika w sposób przypadkowy lała się albo w kierunku szklanki, albo w naszym i w żaden sposób nie moglibyśmy tego przewidzieć; jedno jest pewne: żylibyśmy krócej.



Problem „Dlaczego ruch przebiega właśnie tak, a nie inaczej?” był stawiany od tysiącleci, ale w genialny sposób został rozwiązany dopiero przez Newtona. Jak wiemy ze szkoły, w opisie tym stosujemy pojęcie sił i podstawowe równanie mówi, że przyspieszenie jest proporcjonalne do wypadkowej siły. Jeżeli znamy położenie początkowe, początkową prędkość i siły działające w układzie, to w klasycznym opisie Newtona cały późniejszy ruch jest już konsekwencją tego równania. Ze względu na wagę problemu po Newtonie usiłowano opisać ruch ciał wychodząc również z innych założeń. Sformułowano w tym celu kilka tzw. zasad minimalnych (np. zasadę Maupertuisa czy zasadę Jacobiego).

W tym artykule chcę omówić najważniejszą z nich, tzw. zasadę najmniejszego działania, zwaną również zasadą Hamiltona lub Hamiltona-Jacobiego (zasada ta nosi tę nazwę z powodów historycznych, gdyż w rzeczywistości jest zasadą stacjonarnego działania, a niekoniecznie minimalnego). Zasada ta mówi, że dla rzeczywistych trajektorii pewna wielkość zwana działaniem (której sens postaram się wyjaśnić poniżej) osiąga wartość stacjonarną, tzn. małe, liniowe odchylenie się od tej trajektorii powodowałoby kwadratową (lub jeszcze szybciej zbiegną do zera) zmianę działania. Takie odchylenia są zwykle nazywane wariacjami i dlatego mówimy, że na trajektorii klasycznej znika wariacja działania. Istotne jest, że w sformułowaniu tym nie mówi się nic o siłach i przyspieszeniach, ale można pokazać, że w mechanice klasycznej jest ono (prawie) równoważne opisowi ruchu za pomocą równań Newtona. Niecałkowita równoważność wynika z tego, że inaczej określamy ruch w obydwu przypadkach: w równaniach Newtona określamy punkt początkowy i prędkość początkową (co daje jednoznaczne rozwiązanie), a w zasadzie najmniejszego działania punkt początkowy i końcowy (co może dawać wiele możliwych trajektorii). Poza mechaniką klasyczną okazuje się, że opis dynamiki za pomocą działania jest daleko ogólniejszy niż za pomocą równań Newtona i widać to szczególnie w teoriach fundamentalnych – teorii grawitacji, mechanice kwantowej, kwantowej teorii pola czy najnowszych teoriach, takich jak teoria strun.

Pojęcie działania w mechanice klasycznej można wprowadzić tylko w niektórych sytuacjach, mianowicie wtedy, kiedy w układzie działają tylko tzw. siły zachowawcze, np. nie ma tarcia (typowym przykładem jest ruch planet wokół Słońca). Działanie (oznaczane zwykle przez S) wprowadzamy w następujący sposób: Jeżeli ruch zaczyna się w czasie t_p w punkcie x_p , a kończy w czasie t_k w punkcie x_k , to rozpatrujemy oddzielnie wszystkie możliwe trajektorie w $n + 1$ przedziałach czasu („zdjęciach trajektorii” – patrz rysunek). Dla każdej trajektorii i każdej chwili t_p oraz t_i , $i = 1, \dots, n$, obliczamy w danym punkcie x_i



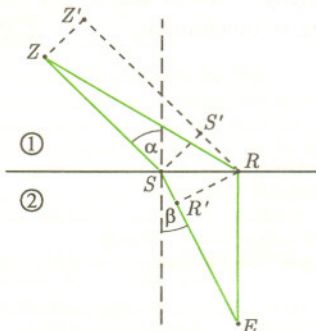
Podział czasu na $n + 1$ przedziałów dla dwóch dowolnie wybranych trajektorii.

Tarcie jest tylko pojęciem makroskopowym, użytecznym wtedy, gdy nie chcemy wnikać, co się dzieje na poziomie atomowym. W teoriach fundamentalnych (mikroskopowych) pojęcie tarcia jest bezużyteczne i wszystkie układy są zachowawcze. Na przykład: hamowanie kulki poruszającej się po stole i „rozpraszanie” jej energii kinetycznej na (niewidoczne mechanicznie) ruchy ciepłe makroskopowo opisujemy jako tarcie, ale mikroskopowo powiemy, że energia kinetyczna kulki zamienia się w trakcie toczenia na energię kinetyczną ruchów ciepłych stołu i kulki. Podobną analizę można przeprowadzić dla wielu innych układów, które wydają się „rozpraszać” energię i na wystarczająco głębokim poziomie opisu następuje jedynie przepływ energii między składnikami, ale układ jest zachowawczy i pojęcie działania może być wprowadzone.



Rozwiązanie zadania F 476.

Niech ZSE będzie drogą rzeczywistą światła, a ZRE drogą porównawczą. Kąty α i β są odpowiednio rzeczywistym kątem padania i załamania. Przez R prowadzimy równoległą do ZS , $Z'S$ i $S'R$ są prostopadłe do ZS , a RR' do SE . Wtedy $\angle S'SR = \alpha$ i $\angle SRR' = \beta$. Stąd i z prawa Snelliusa mamy, że $S'R/SR' = v_1/v_2$. Gdyby więc drogą rzeczywistą światła był odcinek $Z'R$, to w ośrodku 1 przebyłoby ono drogę $S'R$ w takim samym czasie, jak drogę SR' w ośrodku 2, czyli $t(S'R) = t(SR')$ lub $t(Z'R) = t(ZSR') = t(ZS) + t(SR')$. Ponieważ przeciwprostokątna $ZR > Z'R$, więc $t(ZSR') < t(ZR)$. Podobnie $RE > R'E$ i $t(R'E) < t(RE)$.



Dodając te nierówności stronami, stwierdzamy, iż czas przejścia drogi $ZSR'E$, równy $t(ZS) + t(SR') + t(R'E)$, jest zawsze mniejszy od czasu $t(ZR) + t(RE)$ przejścia drogi ZRE .

aktualną energię kinetyczną $T = mv^2/2$ i energię potencjalną $U(x_i)$. Następnie dodajemy wkłady od wszystkich punktów (t_i, x_i) do funkcji $L = T - U$ (tzw. lagranżjanu), mnożąc każdy wkład przez długość odpowiedniego przedziału czasu:

$$(1) \quad S_n = L(t_p, x_p)(t_1 - t_p) + L(t_1, x_1)(t_2 - t_1) + \dots + L(t_n, x_n)(t_k - t_n).$$

Działanie S dla tej trajektorii jest zdefiniowane jako granica S_n dla $n \rightarrow \infty$, czyli S_n dla bardzo drobnych przedziałów (niektórzy Czytelnicy rozpoznają w takim określeniu definicję całki względem czasu z funkcji L). Faktyczną trajektorię chcemy znaleźć z zasady najmniejszego działania: dla niej S prawie się nie zmienia, jeżeli trochę ją zmienimy (ściślej, zmiany S są kwadratowe przy liniowym odchyleniu się od trajektorii rzeczywistej). Dla przykładu obliczmy działanie dla cząstki swobodnej ($U = 0$, czyli dla siły równej zero) z trajektoriami podzielonymi dla uproszczenia tylko na dwa przedziały o równej długości ($t_1 - t_p = t_k - t_1 = \Delta t$). Przekonamy się, że z zasady najmniejszego działania można odtworzyć znany fakt, iż w tej sytuacji ruch jest prostoliniowy i jednostajny. Działanie jest dane wzorem

$$(2) \quad \begin{aligned} S_1 &= L(t_p, x_p)(t_1 - t_p) + L(t_1, x_1)(t_k - t_1) = \\ &= \frac{m(x_1 - x_p)^2}{2\Delta t^2} \Delta t + \frac{m(x_k - x_1)^2}{2\Delta t^2} \Delta t. \end{aligned}$$

Różne x_1 odpowiadają różnym trajektoriom i dają inną wartość działania. Pytanie brzmi: dla jakiego $x_1 = x_{\min}$ działanie S_1 ma minimum? Żeby odpowiedzieć na to pytanie, wystarczy w tym przypadku zauważyć, że S_1 można przekształcić do postaci

$$(3) \quad S_1 = \frac{m(x_1 - (x_p + x_k)/2)^2}{\Delta t} + \frac{m(x_k - x_p)^2}{4\Delta t}.$$

Z postaci tej widać, że minimum działania otrzymujemy dla $x_1 = (x_p + x_k)/2$, co znaczy, że ruch jest prostoliniowy. Również można obliczyć, że prędkość w pierwszym i drugim przedziale jest taka sama i równa prędkości średniej:

$$(4) \quad \frac{x_{\min} - x_p}{\Delta t} = \frac{x_k - x_{\min}}{\Delta t} = \frac{x_k - x_p}{2\Delta t}.$$

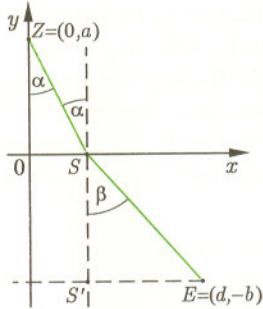
Można wykazać, że dla tego ruchu wnioski te są identyczne dla poprawnie obliczonego działania (czyli granicy S_n przy $n \rightarrow \infty$). Dodajmy, że istnieje znacznie bardziej efektywna metoda obliczania działania poprzez rozwiązywanie równania Hamiltona–Jacobiego i jest ona zresztą najogólniejszym sposobem rozwiązywania problemów w mechanice klasycznej, ale nie będziemy jej tutaj omawiać.

Zasada najmniejszego działania, którą przedstawiliśmy powyżej, obowiązuje w tej formie jedynie w mechanice klasycznej. W mechanice kwantowej zmienia się ona w sposób dość zasadniczy: każda trajektoria jest dozwolona, tylko im dalej jest od trajektorii klasycznej (czyli im większe jest działanie), tym mniejsze jest prawdopodobieństwo jej wystąpienia (czyli trajektoria klasyczna jest najbardziej prawdopodobna, ale już nie jedyna). W mechanice kwantowej dla danych t_p, x_p, t_k i x_k znajdujemy tzw. amplitudę, czyli pewną liczbę, którą możemy obliczyć znając działanie wzdłuż wszystkich trajektorii łączących x_p z x_k (a nie tylko wzdłuż trajektorii klasycznej). Trajektoria klasyczna daje zwykle największy wkład do amplitudy (co jest związane z minimum działania wzdłuż tej trajektorii) i to tłumaczy w pewien sposób zasadę najmniejszego działania w mechanice klasycznej: cząstka „sprawdza” wszystkie trajektorie, ale ta, która minimalizuje działanie, daje największy wkład do amplitudy, więc gdy można cząstce przypisać określoną trajektorię (tak jak dla obiektów makroskopowych w mechanice klasycznej), to ona jest wybierana. Sumowanie wkładów do amplitudy od wszystkich trajektorii nosi nazwę całki po trajektoriach i jest dość trudną procedurą – istnieje jednak prostszy sposób, przez rozwiązywanie pewnego równania różniczkowego zwanego równaniem Schrödingera. Równanie to jest bardzo podobne do równania Hamiltona–Jacobiego z mechaniki klasycznej (co jeszcze raz podkreśla ważność działania), ale zupełnie inna jest jego interpretacja.



Rozwiązanie zadania F 475.

Bierzemy pod uwagę hipotetyczny (niekoniecznie rzeczywisty) promień wysłany z Z pod dowolnym kątem α , $0 < \alpha < \pi/2$, który jest równy kątowi padania w punkcie S (rysunek). Następnie w punkcie tym tak dobieramy kąt załamania β , żeby punktem końcowym promienia załamane go był ustalony w zadaniu punkt E .



Ponieważ $ZS = a / \cos \alpha$ i $SE = b / \cos \beta$, więc całkowity czas, w którym światło przejdzie drogę ZSE , jest równy

$$(1) \quad t = \frac{a}{v_1 \cos \alpha} + \frac{b}{v_2 \cos \beta}.$$

Łatwo możemy sprawdzić, że kąt β jest funkcją kąta α , gdyż z $\triangle OZS$ i $\triangle S'E$ wynika, iż

$$(2) \quad a \tan \alpha + b \tan \beta = d.$$

W rezultacie czas t dany wzorem (1) jest funkcją jednej zmiennej α . Obliczając pochodną t' funkcji t względem α i korzystając z warunku koniecznego istnienia ekstremum, $t' = 0$, dostajemy

$$(3) \quad t' = \frac{a \sin \alpha}{v_1 \cos^2 \alpha} + \frac{b \sin \beta}{v_2 \cos^2 \beta} \beta' = 0,$$

gdzie β' jest pochodną funkcji β względem α . Aby wyznaczyć β' , obliczamy pochodną względem α obu stron równania (2) i stwierdzamy, że

$$(4) \quad \beta' = -\frac{a \cos^2 \beta}{b \cos^2 \alpha}.$$

Eliminując β' z (3) i (4), otrzymujemy prawo Snelliusa

$$\frac{\sin \alpha}{\sin \beta} = \frac{v_1}{v_2}.$$

Dla klasycznej teorii pola, takiej jak klasyczna elektrodynamika czy teoria grawitacji Einsteina, mamy pewne pola zależne zarówno od czasu, jak i od punktu w przestrzeni (takim polem mógłby też być np. rozkład temperatury nad Europą). W tym przypadku „trajektoria” to dowolnie wybrany rozkład pola w czasie i przestrzeni, a działanie dla tego rozkładu jest zdefiniowane jako suma wkładów od wszystkich punktów czasu i przestrzeni dla tej wybranej konfiguracji. W przypadku teorii pola zasada stacjonarnego działania mówi, że rzeczywisty rozkład to taki, dla którego działanie dla „trochę” zmienionego rozkładu pozostaje „prawie” niezmienione (ściślej: liniowe odchylenia od trajektorii powodują jedynie kwadratowe odchylenia działania). Ze względu na to, że nawet dla jednej „trajektorii”, czyli rozkładu pola, bez znajomości pojęcia całki wielokrotnej trudno jest obliczyć odpowiadające jej działanie, trudniej również niż poprzednio podać przykład ilustrujący zasadę stacjonarnego działania. Z zasady tej wynikają pewne równania różniczkowe (równania pola, analogiczne do równań Newtona w mechanice), które często łatwiej jest rozwiązać niż obliczać działanie i znajdować ekstremum. Równania te są jednak równaniami różniczkowymi cząstkowymi i „łatwiej” bardzo rzadko oznacza „łatwo”. Do takich równań pola należą równania Maxwella w elektrodynamice (gdzie działanie to całka z różnicy kwadratów wartości pola elektrycznego i magnetycznego) i równania Einsteina w teorii grawitacji (gdzie działanie to całka ze składowej krzywizny czasoprzestrzeni). Jak wiadomo, znamy rozwiązania tych ostatnich równań jedynie w szczególnych (ale bardzo ważnych) przypadkach.

W kwantowej teorii pola podobnie jak poprzednio musimy sumować po wszystkich „trajektoriach”, czyli rozkładach pola i o ile poprzednio znalezienie amplitudy było bardzo trudne, ale jeszcze jakoś możliwe, to tu sprawa wydaje się beznadziejna. Wrażenie to jest w pewnej mierze słuszne – sprawa nie jest może aż tak beznadziejna, ale amplitudy przejścia od jednej konfiguracji pola do drugiej znamy jedynie w najprostszych przypadkach, jakiegokolwiek bardziej skomplikowane umiemy rozwiązywać jedynie w sposób przybliżony, a są ważne przypadki (np. budowa protonu), o których mimo dziesięcioleci intensywnej pracy niewiele możemy powiedzieć. Wydaje się, że istotny postęp w tej dziedzinie wymaga wypracowania nowych metod również od strony matematyki i jest tu jeszcze bardzo wiele do zrobienia.

Podsumowując: choć pojęcie działania powstało jako jeden z wielu równoważnych opisów ruchu, okazało się później być fundamentalnym sposobem opisu również w najnowszych teoriach (jak teoria strun), gdzie w ogóle definiujemy całą teorię przez podanie działania. Zasada najmniejszego działania jest podstawową zasadą w fizyce klasycznej i pozwala znaleźć dynamikę układu bez odwoływania się do pojęcia siły. W fizyce kwantowej przestaje ona obowiązywać, ale również tutaj działanie jest podstawowym obiektem definiującym teorię i pozwalającym obliczać amplitudy prawdopodobieństwa dla trajektorii.

Wielu uczniów i studentów często również kieruje się „zasadą najmniejszego działania”: „Nie rób tego, czego od ciebie nie wymagają”, ale w szkole czy na studiach daje ona zdecydowanie gorsze rezultaty niż w mechanice.



Rozwiązanie zadania M 846.

Niech S', L', D' będą spodkami wysokości, opuszczonych z wierzchołków S, L, D odpowiednio. Twierdzimy, że trójkąt $S'L'D'$ jest szukanym ROP -em. Niech $\triangle ROP$ będzie dowolnym trójkątem, wpisanym (jak w treści zadania) w $\triangle SLD$. Ponadto, niech R', R'' będą obrazami punktu R przy symetriach osiowych względem prostych LD i SD odpowiednio. Wtedy $|RP| + |PO| + |OR| = |R'P| + |PO| + |OR'| \geq |R'R''|$, przy czym równość zachodzi wtedy i tylko wtedy, gdy $P, O \in R'R''$.

Mamy ponadto $\angle R'DR'' = 2\angle SLD$, $|DR'| = |DR''| = |DR| \geq |DD'|$, więc obliczając podstawę trójkąta $R'DR''$ dostajemy $|R'R''| \geq 2|DD'| \sin(\angle SLD)$ (na rysunku $\alpha = \angle SLD$, $x = DR' \sin \alpha$). Ostatecznie więc $|RP| + |PO| + |OR| \leq 2|DD'| \sin(\angle SLD)$, przy czym równość zachodzi wtedy i tylko wtedy, gdy $R = D'$ i $P, O \in R'R''$.

Dowiedliśmy więc, że dla dokładnie jednego trójkąta minimum jest osiągnięte. Gdyby teraz nasz trójkąt nie był trójkątem $S'L'D'$, to powtarzając rozumowanie powyższe (biorąc zamiast R pozostałe wierzchołki: O lub P) znaleźlibyśmy trójkąt wpisany w $\triangle SLD$ o mniejszym obwodzie.

