Optimalizácia

Víťazná stratégia závisí od cieľa hry

Uvažujme nasledujúci scenár: Detektív sleduje cestujúceho, o ktorom vie, že ide z mesta A do mesta B. Na výber má množstvo spojníc medzi mestami. Detektív sa snaží zistiť kľúč k rozhodovaniu cestujúceho. Existuje niečo, čo sa dotyčný snaží svojimi voľbami **optimalizovať**? Sledovaný človek najprv volí časť najpriamejšej spojnice medzi A, B. Detektív vezme ako predbežnú hypotézu, že sa snaží minimalizovať prejdenú dráhu alebo čas trvania cesty. Na úsekoch, ktoré minimalizujú aj čas aj prejdenú dráhu, detektív nevie rozhodnúť, ktoré krirérium výberu je vyššie. **Kľúčové sú situácie, keď rôzne priority vedú k rôznym výberom.** Pokiaľ vopred nevieme, aké rôzne priority môžu byť v hre, môže byť vážnym problémom vôbec zistiť, ktoré situácie si všímať a kedy sme už s kontrolou hotoví.

Po chvíli sa cestujúci odkloní od priamej spojnice medzi mestami A, B a ide okľukou, a až od istého bodu sa pohybuje opäť priamo k B. Okľuku robil v oblasti, kde sa na priamej línii medzi A a B nachádza akási obec. Detektív vie, že v obci je nutná znížená rýchlosť, a najnovšie aj dopravné obmedzenia pre akési podujatie, ktoré sa tam koná. Keďže cestovateľova obchádzka šetrila čas na úkor šetrenia dráhy, detektív usúdi, že azda sa jedná prioritne o mininalizovanie času. Na základe tohoto predpokladu sa detektív môže pokúsiť predpovedať, ako si bude sledovaný muž cestu voliť ďalej, a overiť svoju teóriu na základe ďalšieho pozorovania.

Môže sa však stať, že predpovede budú detektívovi potvrdzovať jeho aktuálnu teóriu aj napriek tomu, že ešte neuhádol najvšeobecnejší, najvyšší princíp výberu - iba zatiaľ nenarazil na situáciu, v ktorej by tento princíp nariaďoval iné voľby, než aké by boli urobené aj na základe minimalizácie času.

Napríklad na ďalšom úseku ide náš cestujúci poľnými nespevnenými komunikáciami namiesto krajskej cesty. Minimalizácie vzdialenosti ani času by k tejto voľbe neviedli, lebo krajská cesta umožňuje v oboch ohľadoch efektívnejšiu voľbu.

Nakoniec sa môže ukázať, že cestujuci sa usiloval minimalizovať počet stretnutí s policajnými hliadkami - na začiatku išiel čo najpriamejšie, lebo na tom úseku to zároveň šetrilo čas strávený na cestách a pravdepodobnosť že nejakú hliadku stretne; obci sa vyhol, aby jednak znížil túto pravdpodobnosť stretnutia spojenú so samotným zotrvávaním vonku a jednak preto, že počet hliadok tam bol vyšší pre podujatie, čo sa tam chystalo (a ktoré detektív interpretoval ako potvrdenie svojej hypotézy o minimalizácii času). Na poslednom úseku šiel sledovaný muž poľnými cestami celkom bez policajných hliadok, lebo také cesty boli konečne k dispozícii. Okolo jeho štartovacieho mesta neboli, preto tam detektív ani nemal možnosť brať do úvahy možnosti s nimi spojené.

Všimnime si, že detektív sa musel zaoberať komplexným pohľadom na situáciu, aby prišiel k správnemu záveru. Bolo potrebné správne interpretovať, akú rolu hrala absencia poľných ciest blízko mesta A a ich výskyt okolo mesta B, znížená povolená rýchlosť a zvýšená ostraha v obciach medzi mestami, atď. Tým sa dostávame k veľmi dôležitému pojmu - k **väzbám**, obmedzeniam a s tým suvisiacej **optimalizácii v rámci možností**. Nesprávne interpretované pozorovania sa môžu stať zdrojom väčších zmätkov, než aké by nastali bez daných pozorovaní, ale to je samozrejme len upozornenie, nie argument pre ukončenie pozorovaní a pokusov o ich interpretácie.

Obr. 1: Prečo práve tieto cesty? Akú má agendu?

Aj vo vede robíme detektívne úvahy. Sledujeme kroky, ktoými sa systém dostal z A do B, vykresľujúc krivku v abstraktnom či "konkrétnom"priestore (závisiac od toho či A,B predstavujú stav, konfiguráciu,...). Pýtame sa, či existujú aj v prírode vedúce princípy v zmysle, že lokálne kroky sú volené tak, že niečo bolo globálne (v rámci celého deja) **extremalizované**. Opäť v rámci možností, so zohľadnením väzieb. V závislosti od väzieb môžu dva systémy s rovnakou agendou vyzerať veľmi rôzne.

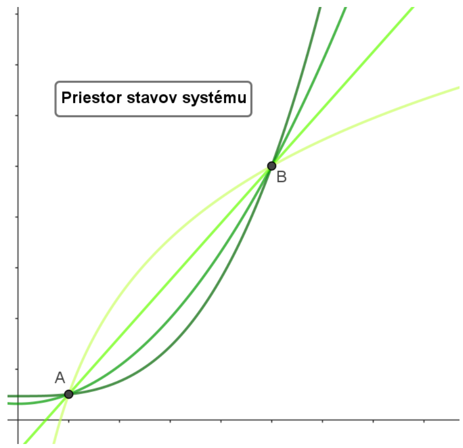
Obr. 2: V islandských vodopádoch aj delte Okavanga má voda tendenciu ísť smerom k nižšiemu gravitačnému potenciálu. Rozdielne väzby majú za následok veľmi rôzne prejavy sledovania tej istej agendy.

V mechanike sme zvyknutí mať stav telesa zadaný polohou x a hybnosťou p, a stav o malý časový interval neskôr bude daný tým aktuálnym a pravidlami hry (pohybovými rovnicami) diktujúcimi nasledujúci krok. Poznať tieto je podobné detektívovej znalosti veľmi lokálnych, krátkodobých rozhodnutí sledovaného človeka.

Ak máme daný stav - bod vo fázovom priestore, Hamiltonove rovnice nám hovoria, ako vyzerá prvá derivácia, teda prvý krok z daného miesta, resp. prvý člen v taylorovskom rozvoji, v závislosti od energie.

$$\dot{x} = \frac{\partial E}{\partial p}$$
$$\dot{p} = -\frac{\partial E}{\partial x}$$

Potom sme v novom stave, (v novom bode fázového priestoru),a môžeme vyhodnotiť Hamiltonove rovnice pre tento stav, teda získať inštrukcie pre ďalší krok atď. Takto sa iteráciami dostávame do ďalších a ďalších stavov, aj takých, ktoré boli neinfitezimálne ďaleko od počiatočného stavu. Dá sa potom skúmať celá množina prejdených stavov z globálneho pohľadu. Je to krivka vo fázovom priestore parametrizovaná časom - každý bod predstavuje stav systému v danom čase. Čím je taká vykreslená krivka - ako celok - význačná medzi všetkými ostatnými ktoré predstavujú nezrealizované postupy priestorom stavov? Vieme matematicky sformulovať túto význačnosť?



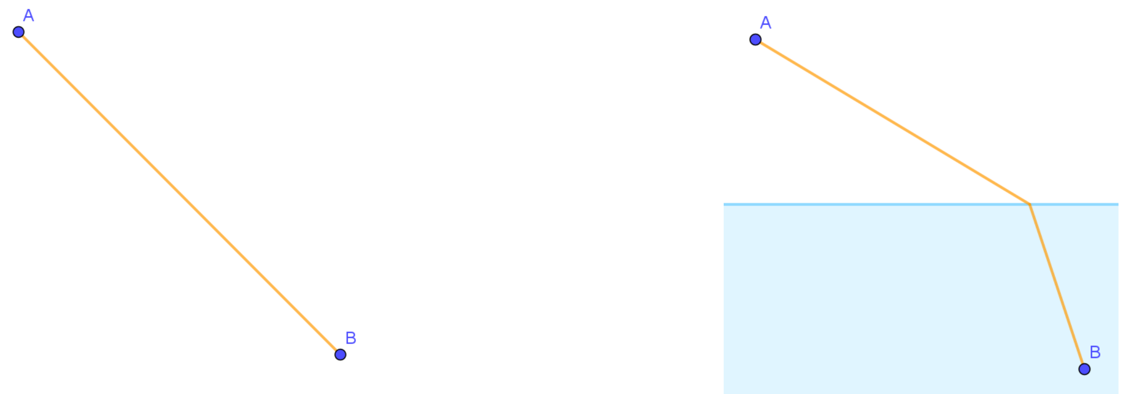
Obr. 3: Podľa čoho si systém vyberie cestu zo stavu A do stavu B?

Váženie ciest

Ako veľmi užitočný nástroj sa tu ukazuje pojem tzv. **funkcionálu**. Umožňuje priradiť čísla objektom ako su napríklad parametrizované krivky vo fázovom prietore.
Teda funkcionál je akási funkcia na funkciách, predpis priraďujúci celej funkcii jedno číslo. Každej možnej krivke spájajúcej dva fixné body vo fázovom priestore priradíme pomocou vhodného fukcionálu číselnú hodnotu, a potom môžeme hladať krivku, ktorej bola priradená extremálna - minimum či maximum. Existuje množstvo funkcionálov definovaných na množine kriviek, a pre rôzne funkcionály máme rôzne krivky ako extremálne, význačné. Fyzik však má kritérium na výber správneho funkcionálu - má to byť ten, podľa ktorého je extremálna krivka práve tá, ktorá sa realizovala. Teda vlastne sa snažíme pomocou funkcionálu a jeho extremalizácie **matematicky reprezentovať výberové kritérium** prírody, aspoň na nejakej úrovni.

Zdá sa, že pre mnohé deje v prírode sa nejaký taký princíp dá nájsť. Funkcionál priraďujúci krivkám v priestore stavov (teda vlastne **časovým vývojom stavov**) číslo voláme **účinok**. Vo fyzike je sa vyskytuje v klasickej aj kvantovej mechanike, v špeciálnej aj všeobecnej relativite. Samozrejme, to, že nám predpovede urobené na základe extremalizácie nejakého účinku vychádzajú, ešte neznamená, že sme ten účinok uhádli správne. Možno sme prišli len na kritérium, ktoré sa s tým ozajstným zhoduje v situáciách, na ktoré sme zatiaľ narazili. Vo fyzike sa teórie nedokazujú, len sa porovnávajú s pozorovaním a ich kredibilita stúpa, ak s ním súhlasia. Je potrebné robiť ďalšie pozorovania v rozličných situáciách a ak sa v niektorej náš účinok ukáže ako nepostačujúce kritérium (systém sa nespráva tak aby ho extremalizoval), pokúšame sa nájsť vyšší princíp než ten, čo práve zlyhal.

Príkladom je sledovanie svetelných lúčov v klasickej geometrickej optike. V rovnorodom prostredí možno prídeme k hypotéze, že lúč si vyberá cestu tak, aby minimalizoval prejdenú dráhu. Alebo čas - tu to vedie k rovnakej stratégii. Potom však vyskúšame šírenie v aspon dvoch prostrediach, v ktorých sa svetlo šíri s rôznou rýchlosťou, a zistíme, že šetrenie geometrickej dráhy nevyzerá byť najvyššia priorita, šetrenie času je uprednostnené. Podobne ako detektív z podobenstva na začiatku, nemôžeme ešte prehlásiť, že už o systéme vieme všetko[1].



Obr. 4: Vo vode sa signál šíri pomalšie. Pre šetrenie času sa oplatí "nadbehnúť"si časť vzdialenosti na vzduchu, aj za cenu predĺženia geometrickej dráhy.

Silné postavenie, ktoré má účinok vo fyzike, je naznačené jednak tým, že sa vyskytuje v toľkých fyzikálnych modeloch, ale aj v tom že jeho **fyzikálnou jednotkou je J.s**, jednotka v akej je aj **Planckova konštanta**. Pripomeňme tu, že táto je jednou z fundamentálnych konštánt v prírode, ktorej hodnotu zatiaľ nevieme z ničoho odvodiť (len namerať). Je spojená s mnohými javmi v prírode. Veličiny merané v jednotkách ako ona môžu mať význam hlbší ako tie, ktoré sme zaradili v rámci SI sústavy ako "základné"(lebo pre bytosti s našimi rozmermi, zmyslami a spôsobom vyjadrovania sa tak javili). Táto myšlienka je minimálne podporená veľkým významom **momentu hybnosti** naprieč fyzikou, ktorý je tiež meraný v spomenutých jednotkách. Pre ľudí používajúcich mechaniku predovšetkým na technické výpočty je to azda jedna z mnohých odvodených veličín, ale ak si uvedomíme, že kvantový spin častíc s celým jeho dopadom na správanie hmoty je tiež druhom momentu hybnosti, jeho osobitné postavenie začína byť výraznejšie, ak aj nie jasnejšie.

Optimalizáciou funkcionálov (hľadaním extrémov pre ne) sa v matematike zaoberá **variačný počet**. Proces varírovania funkcionálu, teda hľadanie funkcií (alebo ich grafov - kriviek) ktoré ho extremalizujú, sa trochu podobá hľadaniu extrému funkcie, teda hľadaniu bodov v ktorých je extremálna. Užitočnosť derivácie ťažko poprieť - okrem iného ako nástroja umožňujúceho hľadať extrémy. Na základe toho možno aspoň trochu odhadnúť význam funkcionálu a hľadania jeho extrémov, keďže sa vlastne jedná o spôsob, ako celej **funkcii priradiť číslo - akúsi váhu** a na jeho základe urobiť "hierarchiu"medzi funkciami, a nájsť tie, ktoré sú v rámci nej najvýznačnejšie. Sú to veľmi silné nástroje nielen vo fyzike. Matematicky sformulovateľné úlohy z mnohých oblastí, kde je potrebné robiť optimalizáciu niečoho, sú často riešiteľné práve tým, že sa identifikuje (toto nemusí byť triviálna úloha, ale je aj akýmsi testom vhľadu do problému) veličina definovaná na jednotlivých možnostiach, ktorá sa má optimalizovať. Správna možnosť je potom pravdepodobne jedna z tých, pre ktoré hľadaná veličina nadobúda extrém.

Štandardný postup pri variačných problémoch môže byť nasledovný: máme funkcionál a množinu funkcií na ktorých je definovaný - varírovanám nájdeme (tzv. **Euler-Lagrangeove**) rovnice ktorých **riešenia môžu byť význačné funkcie**. Upozornime tu, že byť riešením takýchto rovníc je pre funkciu len nutnou, nie postačujúcou podmienkou extremalizácie funkcionálu. Veľmi podobne to bolo v obyčajnej analýze - tak splnenie $\frac{df}{dx} = 0$ v nejakom bode x bolo nutné ak v x bolo minimum alebo maximum, ale nebolo to postačujúce. Niekedy máme inverzný problém - máme k dispozícii diferenciálne rovnice popisujúce správanie systému, a pýtame sa či tieto rovnice sú Euler-Lagrangeove pre nejaký funkcionál. Nie je to len akademická otázka - nakoniec poznať funkcionál znamená do istej miery **identifikovať globálnu agendu**, ktorú systém sleduje.

[1]Už trocha kvantovej mechaniky a relativity nám povie že nevieme.

Aj newtonovská mechanika zahŕňa optimalizáciu

Ako príklad spomeňme newtonovskú mechaniku. Mnohokrát sa študenti najprv naučia Newtonove pohybové rovnice, obvykle bez úvah o tom, či systém ich poslúchaním optimalizuje nejakú veličinu. **Z matematického hľadiska je druhý zákon diferenciálna rovnica druhého rádu, Euler-Lagrangeova rovnica, pre funkcionál predstavujúci rozdiel kinetickej a potenciálnej energie integrovaný počas celého časového vývoja vývoja.** Ilustrujme si tu čo to znamená na jednoduchom príklade.

Najprv zvážme voľnú časticu pohybujúcu sa pozdĺž osi x. Nech v čase t_1 bola v mieste X_1 a v čase t_2 v mieste X_2. Jej trajektória nech je nejaká časom parametrizovaná krivka - úsek na osi x, vo zvolenej vzťažnej sústave povedzme $x(t)$, pričom $x(t_1) = X_1$, $x(t_2) = X_2$. Jedná sa o voľnú časticu, potenciálna energia je teda nulová a globálnou agendou nášho telesa bude extremalizovať len kinetickú energiu preintegrovanú cez celý čas pohybu.

$$\int_{t_1}^{t_2} dt \frac{1}{2} m \dot{x}_i \dot{x}_i$$

Ako sa bude častica pohybovať v čase $t_1 < t < t_2$ medzi bodmi X_1 a X_2 ? Človek ani nemusí vedieť, ako sa hľadá extrém funkcionálu, aby vedel, že v tomto prípade častica pôjde stále rovnako veľkou rýchlosťou na základe nasledujúcej úvahy: ak má za čas $t_2 - t_1$ prejsť vzdialenosť $X_2 - X_1$, potom priemerná rýchlosť musí byť

$$v_p = \frac{X_2 - X_1}{t_2 - t_1}$$

Uvažujme najprv pre jednoduchosť, že ide stále rýchlosťou v_p, okrem časového intervalu dt kde ide rýchlosťou $v_p - \epsilon$, a nutne aj časového intevalu dt keď musí za stratu kompenzovať a ísť rýchlosťou $v_p - \epsilon$. Zložitejšie fluktuácie rýchlosti sa dajú vyskladať analogicky. Keby výraz pod integrálom bol úmerný len prvej mocnine rýchlosti, fluktuácie by sa kompenzovali a hodnota integrálu sa nezmenila. Ale v prípade druhej mocniny rýchlosti kompenzácia nie je úplná a akákoľvek fluktuácia (za podmienky, že teleso prejde za daný čas danú dráhu) hodnotu integrálu navýši: uvážme príspevky z úseku kde teleso ide o ϵ pomalšie a z úseku kde ide o ϵ rýchlejšie. Prispievajú hodnotami

$$\frac{1}{2} m (v_p - \epsilon)^2 dt = \frac{1}{2} m (v_p^2 - 2\epsilon v + \epsilon^2) dt$$

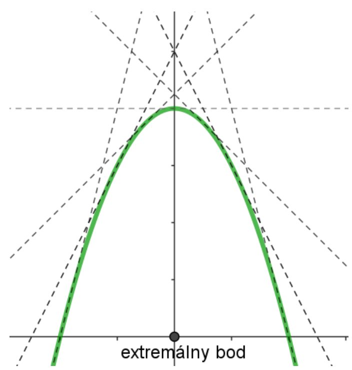
$$\frac{1}{2} m (v_p + \epsilon)^2 dt = \frac{1}{2} m (v_p^2 + 2\epsilon v + \epsilon^2) dt$$

Vidíme, že pri sčítaní vypadnú členy úmerné prvej mocnine rýchlostnej fluktuácie, ale tie úmerné druhej mocnine sa sčítajú. Celkovo teda do integrálu oproti rovnomernému pohybu pribudne kladný príspevok $m\epsilon^2 dt$.
Teda na základe samotnej informácie, že systém (tu: voľná častica) má agendu extremalizovať istý určitý integrál, vieme nanovo odvodiť pravidlo pre rovnomerný priamočiary pohyb pre voľnú časticu - zatiaľ aj bez znalosti ako extremalizovať všeobecné funkcionály!

Rozvíjanie nástrojov: Funkcionálna derivácia

Spomeňme si, ako sa hľadali extrémy funkcie. Funkcia je predpis, ktorý každému bodu z definičného oboru priradil nejaký bod z oboru hodnôt. Body definičného oboru boli obvykle buď čísla alebo usporiadané skupiny čísel(pre funkcie viacerých premenných). Body oboru hodnôt boli čísla, v prípade vektorových funkcií usporiadané skupiny čísel (ale vektorová funkcia sa dá chápať ako usporiadaný súbor skalárnych funkcií, tak tu nebudeme ďalej spomínať takéto všeobecnejšie prípady.) Pri hľadaní extrému sa vlastne hľadal taký bod z definičného oboru, ktorému funkcia priradila nejaké lokálne výnimočné číslo - buď menšie alebo väčšie ako všetkým ostatným bodom v jeho okolí. Pre spojité funkcie navyše platilo, že všetky body v okolí toho výnimočného majú funkčnú hodnotu odlišnú až v druhom ráde.

Ilustrujme myšlienku na príklade funkcie jednej premennej, ktorej graf nech predstavuje kopec. Povedzme že vieme robiť nekonečne malé kroky. Pre body na svahu platí, že každým krokom sa posunieme buď hore alebo dole. Strmosť stúpania či klesania je daná dotyčnicou v danom bode, ktorej smernica je práve prvá derivácia v tom bode. Prvá derivácia je zároveň faktor násobiaci veľkosť infinitezimálneho kroku v taylorovskom rozvoji aproximujúcom funkciu. V bode na vrchole kopca je dotyčnica nulová a smernica vodorovná. Do prvého rádu (vo veľkosti kroku) sú hodnoty funkcie v okolí maxima rovnaké. Samozrejme v ďalších rádoch sa už líšia, inak by nešlo o lokálny extrém. Pri hľadaní extrému sa poobzeráme po hodnotách funkcie v okoliach skúmaných bodov



Obr. 5: Do prvého rádu sa hodnoty funkcie v okolí maxima nemenia - funkcia nestúpa a neklesá.

a ako z extrému podozrivé body identifikujeme tie, v okolé ktorých sa funkcia do prvého rádu nemení. (Až potom skúmame druhú deriváciu, aby sme rozlíšili inflexný a extremálny bod.)

Analogicky postupujeme pri **hľadaní extémov pre funkcionály**. Zhruba povedané, funkcionál je predpis, ktorý **priraďuje čísla funkciám**. Z geometrického hľadiska, prvky ("body"v abstraktnom zmysle) jeho definičného obor sú krivky (grafy funkcií), a jeho obor hodnôt sú čísla. Priradiť číslo funkcii sa dá všelijako. Napríklad by sme každej mohli priradiť priemer jej hodnôt. Alebo jej maximálnu alebo minimálnu hodnotu, pre funkcie pre ktoré to má zmysel. A tak ďalej - podobne ako existuje mnoho funkcií (spôsobov ako číslu priradiť číslo), je aj mnoho funkcionálov (spôsobov ako funkciám priradiť číslo). Jedným z veľmi užitočných funkcionálov je určitý integrál nejako závislý na našej funkcii. Veľmi často aj na jej deriváciách. Naozaj, určitý integrál je spôsob ako z funkcií dostávať čísla:

$$\mathcal{S}: \; f \mapsto \mathcal{S}[f] = \int_A^B dx\, L(x, f(x), f_x(x), f_{xx}(x), ...) \tag{1}$$

Výrazu L pod integrálom sa hovorí **Lagranžián**. Predstavuje nejakú kombináciu nezávislých a závislých premenných a ich derivácií[2]. V jeho konkrétnej podobe je kódovaná agenda systému, teda *cieľ hry* pri optimalizácii. Samozrejme nie každá funkcia taký integrál vôbec má, ale to je v poriadku. Funkcionál je definovaný len pre istý definičný obor funkcií, pre ktoré taká operácia má zmysel (podobne ako funkcie boli definované len pre istý definičný obor číslených hodnôt). Podobne ako sme sa pri funkciách mohli pýtať, pre ktorý prvok definičného oboru nadobúdajú maximum či minimum, sa môžeme pre funkcionály pýtať, pre ktorý prvok ich definičného oboru (teda pre akú funkciu) nadobúdajú extrém. Pri hľadaní použijeme analóg derivácie - **funkcionálnu**

[2] Tu f_x, f_{xx},... značia prvú, druhú,... deriváciu podľa premennej. Začíname s funkciou jednej premennej, ale variačný počet sa samozrejme dá - vcelku triviálne - zovšeobecniť na prípady viac závislých aj nezávislých premenných.

deriváciu. *Z extremálnosti sú podozrivé tie funkcie, v okolí ktorých sa funkcionál do prvého rádu nemení.* Nezabudnime, že okolím sa myslia blízke prvky definičhého oboru, čo sú tu funkcie. Pri obyčajných deriváciách sme skúmali, ako sa malé kroky v definičnom obore prejavia na funkčných hodnotách:

$$x \to x + \epsilon \qquad\qquad f(x) \to f(x+\epsilon) \approx f(x) + \epsilon \frac{df}{dx}$$

Pri funkcionálnych deriváciách skúmame, ako sa malé kroky v definičnom obore prejavia na hodnote funkcionálu:

$$f(x) \to f(x) + \epsilon\eta(x) \qquad\qquad \mathcal{S}[f] \to \mathcal{S}[f+\epsilon\eta] \approx \mathcal{S}[f] + \epsilon \frac{\delta \mathcal{S}}{\delta f} \tag{2}$$

Tu $\eta(x)$ je **ľubovoľná variácia** či modifikácia k skúmanej funkcii f, **nulová v okrajivých bodoch** intervalu, cez ktorý integrujeme, teda

$$\eta(x_A) = \eta(x_B) = 0 \tag{3}$$

Výrazu $\frac{\delta \mathcal{S}}{\delta f}$ hovoríme funkcionálna derivácia, keďže naozaj je mierou citlivosti funkcionálu na malú zmenu funkcie (reprezentovanú príspevkom $\epsilon\eta(x)$). Konkrétne vyjadrenie získame jednoduchým rozpísaním $\mathcal{S}[f+\epsilon\eta]$, rozvinutím do prvého rádu v ϵ, použitím základnej vety integrálneho počtu a nulovosti variácie v okrajových bodoch.

$$\mathcal{S}[f+\epsilon\eta] = \int_{x_A}^{x_B} dx\, L(x,\, f+\epsilon\eta, f_x + \epsilon\eta_x,\, f_{xx} + \epsilon\eta_{xx},\, ...) \tag{4}$$

$$\mathcal{S}[f] + \epsilon \frac{\delta \mathcal{S}}{\delta f} = \int_{x_A}^{x_B} dx\, L(x,\, f,\, f_x,\, f_{xx},\, ...) + \epsilon \int_{x_A}^{x_B} dx\, \frac{dL}{d\epsilon}$$

Ak je funkcionálna derivácia pre nejaký argument f nulová, funkcionál $\mathcal{S}$ sa pri zmene argumentu o $\epsilon\eta(x)$ do prvého rádu v ϵ nezmení.

$$\delta \mathcal{S} = 0 \quad \to \quad \int_{x_A}^{x_B} dx\, \frac{dL}{d\epsilon} = 0 \tag{5}$$

$$0 = \int_{x_A}^{x_B} dx\, \frac{dL}{d\epsilon} \; = \int_{x_A}^{x_B} dx\, \left(\frac{\partial L}{\partial f}\eta \, + \, \frac{\partial L}{\partial f_x}\eta_x \, + \, \frac{\partial L}{\partial f_{xx}}\eta_{xx} \, + \, ... \right) \tag{6}$$

Chceme nájsť takú f, aby funkcionálna derivácia bola nulová pre hocijaké η. Potrebujeme podmienku (6) dostať do tvaru

$$0 = \int_{x_A}^{x_B} dx\, \frac{dL}{d\epsilon} = \int_{x_A}^{x_B} dx\, (...)\eta$$

Treba sa teda zbaviť derivácií η_x, η_{xx} atď. Použijeme na to metódu integrovania per partes, základnú vetu integrálneho počtu a fakt, že $\eta(x_A) = \eta(x_B) = 0$. Najprv si to ukážme na člene úmernom η_x. Symbolom D_x označujme totálnu deriváciu podľa nezávislej premennej x.

$$\int_{x_A}^{x_B} dx\, \frac{\partial L}{\partial f_x}\eta_x = \int_{x_A}^{x_B} dx\, D_x\left(\frac{\partial L}{\partial f_x}\eta\right) - \int_{x_A}^{x_B} dx\, \left(D_x \frac{\partial L}{\partial f_x}\right)\eta$$

$$= \left(\frac{\partial L}{\partial f_x}\eta\right)\Bigg|_{x_A}^{x_B} - \int_{x_A}^{x_B} dx\, \left(D_x \frac{\partial L}{\partial f_x}\right)\eta$$

$$= -\int_{x_A}^{x_B} dx\, \left(D_x \frac{\partial L}{\partial f_x}\right)\eta$$

Člen úmerný η_{xx} ošetríme podobne, ale tu samozrejme treba použiť integráciu per partes dvakrát.

$$\int_{x_A}^{x_B} dx\, \frac{\partial L}{\partial f_{xx}}\eta_{xx} = \int_{x_A}^{x_B} dx\, D_x\left(\frac{\partial L}{\partial f_{xx}}\eta_x\right) - \int_{x_A}^{x_B} dx\, \left(D_x \frac{\partial L}{\partial f_{xx}}\right)\eta_x$$

$$= \int_{x_A}^{x_B} dx\, D_x\left(\frac{\partial L}{\partial f_{xx}}\eta_x\right) - \int_{x_A}^{x_B} dx\, D_x\left[\left(D_x \frac{\partial L}{\partial f_{xx}}\right)\eta\right] + \int_{x_A}^{x_B} dx\, \left(-D_x(-D_x)\frac{\partial L}{\partial f_{xx}}\right)\eta$$

$$= \left(\frac{\partial L}{\partial f_{xx}}\eta\right)\Bigg|_{x_A}^{x_B} - D_x\left[\left(D_x \frac{\partial L}{\partial f_{xx}}\right)\eta\right]\Bigg|_{x_A}^{x_B} + \int_{x_A}^{x_B} dx\, \left((-D_x)(-D_x)\frac{\partial L}{\partial f_{xx}}\right)\eta$$

$$= \int_{x_A}^{x_B} dx\, \left((-D_x)(-D_x)\frac{\partial L}{\partial f_{xx}}\right)\eta$$

Vidno, že členy zahŕňajúce vyššie derivácie sa dajú upraviť opakovaným použitím metódy per partes, že nulovosť variácie v okrajových bodoch nás zbaví všetkých členov okrem jedného, ktorý bude mať alternujúce znamienko - plus pre párny počet použitých per partes a mínus pre nepárny.

$$\int_{x_A}^{x_B} dx\, \frac{\partial L}{\partial f_J}\eta_J = \int_{x_A}^{x_B} dx\, \left((-D)_J \frac{\partial L}{\partial f_J}\right)\eta$$

Tu J používame ako multiindex označujúci podľa čoho všetkého sa derivuje a koľkokrát. Pre jednu nezávislú premennú stojí J namiesto x, xx, xxx .... pre viac nezávislých premenných to môže byť trebárs xy označujúce deriváciu podľa prvej a druhej nezávislej premennej atď. Napríklad pre $J = xy$ budeme výraz $(-D)_{xy}$ rozumieť ako $(-D_x)(-D_y) = D_x D_y$.

Teraz sa vráťme k hľadaniu funkcie extremalizujúcej funkcionál (1), teda k požiadavke nulovosti funkcionálnej derivácie (5).

$$\delta\mathcal{S} = 0 \quad \rightarrow \quad \int_{x_A}^{x_B} dx\, \frac{dL}{d\epsilon} = 0 \tag{7}$$

$$\rightarrow \quad \int_{x_A}^{x_B} dx \sum_J \left((-D)_J \frac{\partial L}{\partial f_J}\right)\eta \;=\; 0$$

Suma cez multiindex J nie je nekonečná - nenulové sú len členy zahŕňajúce derivácie f_J vyskytujúce sa v Lagranžiáne L. Člen $J = 0$ je v sume zahrnutý tiež, pričom $D_0 = 1$. Keďže nulovosť (7) má platiť pre ľubovoľnú η spĺňajúcu (3), musí sa nule rovnať faktor, ktorý ju násobí, teda

$$\delta\mathcal{S} = 0 \quad \rightarrow \quad \sum_J \left((-D)_J \frac{\partial L}{\partial f_J}\right) \;=\; 0 \tag{8}$$

Týmto dostávame **podmienku stacionárnosti** funkcionálu $\mathcal{S}[f]$. Je to **nutná**, nie postačujúca podmienka extremálnosti. Na funkcii f, ktorá ju spĺňa, môže mať funkcionál extrém (maximum alebo minimum). Alebo len akýsi analóg inflexného bodu.

Podmienke (8) sa hovorí **Euler-Lagrangeova rovnica**. Ak Lagranžián zahŕňa **viac nezávislých premenných** $x^i,\ i = 1, ...p$, potom multiindex J prebieha cez všetky možnosti relevantných derivácií podľa nich. Ak Lagranžián obsahuje **viac závislých premenných**, povedzme $f^\alpha,\ \alpha = 1, ...q$ potom máme toľko Euler-Lagrangeových rovníc koľko je týchto závislých premenných, a pre stacionárnosť funkcionálu je nutné aby boli splnené všetky, teda aby platilo

$$\sum_J \left((-D)_J \frac{\partial L}{\partial f^\alpha_J} \right) = 0 \qquad \forall \alpha \tag{9}$$

Skúšanie nástrojov: Voľná častica v 1D

Vráťme sa k prípadu voľnej častice. Spomínali sme už, že zákony klasickej mechaniky sa dajú napísať aj vo forme variačného princípu - častice budú sledovať také dráhy v priestore stavov (v klasickej mechanike hmotných bodov je to priestor polôh a hybností), aby extremalizovali rozdiel kinetickej (T) a potenciálnej (U) energie preintegrovaný cez čas pohybu. V prípade voľnej častice ($U = 0$) hmotnosti m v jednom rozmere máme teda nezávislú premennú čas t (okrajové okamihy intervalu označme t_A, t_B), ako závislú premennú máme polohu[3] častice $x(t)$, Lagranžián je jednoducho kinetická energia, a účinok je časový integrál z nej:

$$\mathcal{S}[x] = \int_{t_A}^{t_B} dt\, \frac{1}{2} m x_t^2 = \int_{t_A}^{t_B} dt\ \underbrace{\frac{1}{2} m \dot{x}^2}_{L(t,x,\dot{x})}$$

Podľa dávnej newtonovskej tradície sme časové derivácie označili bodkou. Hľadáme $x(t)$ pre ktorú je tento funkcionál stacionárny, tentoraz už použitím našej mašinérie Euler-Lagrangeových rovníc. Keďže Lagranžián obsahuje len závislú premennú x, jednu nezávislú premennú t, a nanajvýš prvú deriváciu podľa nej, bude pre multiindex J v sume (9) stačiť prebiehať hodnoty 0, t. Navyše náš Lagranžián neobsahuje explicitne x, takže člen zodpovedajúci nulovému J aj tak vypadne.

$$\begin{aligned} 0 = \sum_{J=0}^{t} \left((-D)_J \frac{\partial L}{\partial x_J} \right) &= D_0 \frac{\partial L}{\partial x} - D_t \frac{\partial L}{\partial x_t} \\ &= \left(\frac{\partial}{\partial x} - \frac{d}{dt} \frac{\partial}{\partial \dot{x}} \right) \left(\frac{1}{2} m \dot{x}^2 \right) \\ &= 0 - m\ddot{x} \end{aligned}$$

Inak povedané, voľná častica pôjde bez zrýchlenia, výsledok ktorý poznáme z dávnych štúdií i nedávnych úvah na začiatku kapitoly. Teraz sme tento jav použili ako test nášho nového nástroja, Euler Lagrangoevých rovníc.

Skúšanie nástrojov: Častica v 3D, v poli konzervatívnej sily

Aby sme vyskúšali rovnice (9) pre **viac závislých premenných** a zároveň zovšeobecnili skúmaný systém aj na prípad nenulového potenciálu, vezmime časticu hmotnosti m v trojrozmernom priestore, ktorej poloha je v každom okamihu daná trojicou súradníc $x(t), y(t), z(t)$. Nech potenciálna energia častice v závislosti od miesta je popísaná skalárnou funkciou $U(x, y, z)$. Nezávislá

[3] Zvyk automaticky považovať veci označené ako x za nezávislé premenné má za následok mnoho nedorozumení pri prebiehaní z matematickej literatúry k fyzikálnej. Treba čítať kontext, nie písmená, inak sa podobáme deťom ktoré si neporadia s Pythagorovou vetou ak odvesny a prepona nie sú označené v abecednom poradí.

premenná je čas t (skúmajme pohyb medzi okamihmi označenými t_A, t_B), závislé premenné sú x, y, z. Hľadáme také funkcie $x(t), y(t), z(t)$, pre ktoré je stacionárny funkcionál

$$\mathcal{S}[x,y,z] = \int_{t_A}^{t_B} dt \, \underbrace{\left(\frac{1}{2}m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - U(x,y,z)\right)}_{L(t,x,y,z,\dot{x},\dot{y},\dot{z})}$$

Podmienka stacionárnosti je v tomto prípade vyjadrená tromi Euler-Lagrangeovými rovnicami:

$$0 = \left(\frac{\partial}{\partial x} - \frac{d}{dt}\frac{\partial}{\partial \dot{x}}\right)\left(\frac{1}{2}m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - U(x,y,z)\right) = -\frac{\partial U}{\partial x} - m\ddot{x}$$

$$0 = \left(\frac{\partial}{\partial y} - \frac{d}{dt}\frac{\partial}{\partial \dot{y}}\right)\left(\frac{1}{2}m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - U(x,y,z)\right) = -\frac{\partial U}{\partial y} - m\ddot{y}$$

$$0 = \left(\frac{\partial}{\partial z} - \frac{d}{dt}\frac{\partial}{\partial \dot{z}}\right)\left(\frac{1}{2}m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - U(x,y,z)\right) = -\frac{\partial U}{\partial z} - m\ddot{z}$$

Alebo stručnejšie môžeme zbaliť tri rovnice do vektorového tvaru

$$m\ddot{\vec{r}} = -\vec{\nabla}U$$

Čiže sila je záporne vzatý gradient potenciálnej energie. Pri známej U je známy aj jej gradient, teda jedná sa potom už "len" o obyčajnú diferenciálnu rovnicu (v nezávislej premennej t) druhého rádu a možnosti jej doriešenia (a nájdenia funkcií $x(t), y(t), z(t)$ extremalizujúcich daný funkcionál) značne závisia od konkrétneho tvaru U. Neprišli sme k novinke, ale to nebolo cieľom. Overili sme, že zatiaľ naša variačná metóda dáva výsledky konzistentné s tým, čo sa v mechanike osvedčilo doposiaľ.

Skúšanie nástrojov: Harmonické funkcie ako riešenia variačného problému

Už sme si vyskúšali prípad viacerých závislých premenných. Pre zmenu si teraz metódu ukážme na prípade jednej závislej premennej (volajme ju u) a **viacerých nezávislých premenných** (volajme ich x, y, z)[4]. Nech oblasť integrovania je nejaký objem Ω. Nájdime $u(x,y,z)$ v ktorom je stacionárny funkcionál

$$\mathcal{S}[u] = \int_\Omega dx\,dy\,dz\,\frac{1}{2}(\vec{\nabla}u.\vec{\nabla}u) = \int_\Omega dx\,dy\,dz\,\underbrace{\frac{1}{2}(u_x^2 + u_y^2 + u_z^2)}_{L(x,y,z,u,u_x,u_y,u_z)}$$

Príslušná Euler Lagrangeova rovnica (jedna, lebo máme len jednu závislú premennú) bude

$$0 = \left(\frac{\partial}{\partial u} - \frac{d}{dx}\frac{\partial}{\partial u_x} - \frac{d}{dy}\frac{\partial}{\partial u_y} - \frac{d}{dz}\frac{\partial}{\partial u_z}\right)\left(\frac{1}{2}(u_x^2 + u_y^2 + u_z^2)\right)$$

$$= u_{xx} + u_{yy} + u_{zz} = \Delta u$$

Toto je dobre známa Laplaceova rovnica, tu vystupujúca ako podmienka pre extremalizáciu istého funkcionálu. Na jej riešeniach, harmonických funkciách, je tento funkcionál stacionárny.

[4]Toto je tréning postrehu, pochopenia a ešte niekoľkých ďalších kvalít. V predošlom príklade boli tieto symboly použité na označenie závislých premenných!

Skúšanie nástrojov: Vyššie derivácie

Doposiaľ sme si ukázali variačné metódy na prípadoch, keď Lagranžián obsahoval nanajvýš prvé derivácie, a v tom prípade sú Euler- Lagrangeove rovnice nanajvýš druhého rádu. Takéto majú zatiaľ asi najširšie uplatnenie vo fyzike aj technike. Význam druhej derivácie súvisí s hlbokými geometrickými skutočnosťami, súvisiacimi so samotným priestorom a časom. Ale je dobré vedieť zaobchádzať aj s vyššími deriváciami. Ukážme si teraz našu metódu na prípade nasledovného hypotetického funkcionálu: Nezávislé premenné nech sú x, t, a uvažujem ich na nejakej oblasti Ω. Závislá premenná nech je $u(x,t)$ a povedzme, že agendou systému je extremalizovať funkcionál

$$\mathcal{S}[u] = \int_\Omega dx\, dt \qquad \underbrace{u_{xt}^2}_{L(x,t,u,u_x,u_t,u_{xx},u_{tt},u_{xt})}$$

Suma operátorov v Euler-Lagrangeovej rovnici teda musí siahať ďalej ako v predošlých prípadoch, ale vzhľadom na jednoduchosť Lagranžiánu bude nenulový len jeden člen:

$$\begin{aligned} 0 &= \left(\frac{\partial}{\partial u} - \frac{d}{dt}\frac{\partial}{\partial u_t} - \frac{d}{dx}\frac{\partial}{\partial u_x} + \frac{d^2}{dt^2}\frac{\partial}{\partial u_{tt}} + \frac{d^2}{dx^2}\frac{\partial}{\partial u_{xx}} + \frac{d^2}{dt\, dx}\frac{\partial}{\partial u_{xt}} \right)(u_{xt}^2) \\ &= \frac{d^2}{dt\, dx}\frac{\partial (u_{xt}^2)}{\partial u_{xt}} = 2u_{xxtt} \end{aligned}$$

Riešením je napríklad $u = axt + bx + ct + d$, kde a, b, c, d sú konštanty.

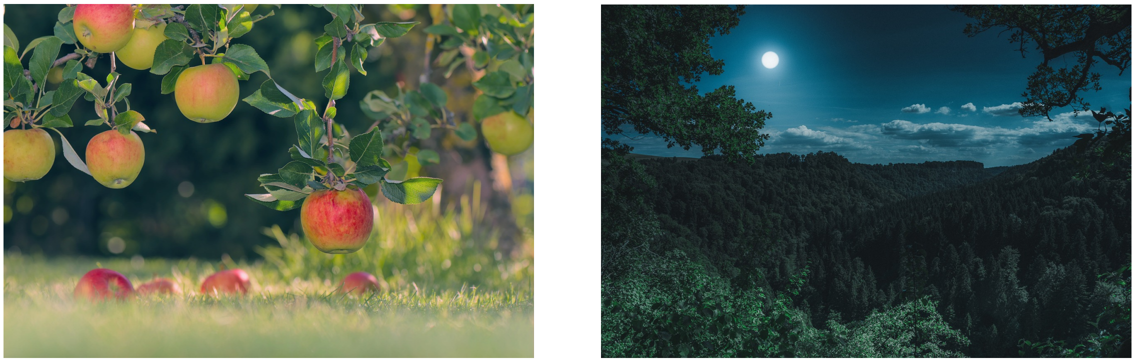
Spoločná agenda a jednotiace princípy

Matematické metódy variačného počtu nám ponúkajú nástroje, ako efektívne hľadať funkcie extremalizujúce rôzne funkcionály. Azda presnejšie je povedať, pomáhajú nám hľadať rovnice, ktoré tieto funkcie majú spĺňať[5].

Ale je tu hlbšia otázka ako len zefektívnenie algoritmu na výpočty. *Ako vieme, čo bude príroda optimalizovať?* Napríklad v prípade klasického pohybu častice, prečo je agendou práve minimalizácia rozdielu medzi kinetickou a potenciálnou energiou? Odpovedať sa dá na viacerých úrovniach a povedzme vopred, ešte je v tých úrovniach čo dopĺňať. Ak takto (ako rozdiel kinetickej a potenciálnej energie preintegrovaný cez čas) zvolíme účinok, ako Euler-Lagrangeove rovnice dostávame práve Newtonov druhý zákon (teraz sa obmedzujeme na prípady keď má zmysel hovoriť o potenciálnej energii, teda o konzervatívnej sile). Aj Newtonove rovnice boli v teórii postulované, nie odvádzané. Uhádnutím správneho účinku (alebo zatiaľ nie preukázateľne nesprávneho) sme v situácii detektíva, ktorému sa podarilo sformulovať kritérium, podľa ktorého si sledovaný jedinec vyberal postup. Samozrejme to nezodpovedá všetky otázky, ktoré môžu byť položené, ale môže to pomôcť robiť predpovede pre ďalšie pozorovania. Otázku prečo má byť optimalizované práve to a to však rozhodne netreba zamietnuť do kúta. Nakoniec optimalizovaná veličina bude zrejme nejakým spôsobom hovoriť veľa. Na tejto úrovni možno ešte nie je dobre vidieť, čo je také hlboké za rozdielom kinetickej a potenciálnej energie nasčítanom na časovom intervale. Z pohľadu štatistickej fyziky môže ísť o akúsi variantu ekvipartičnosti, princípu rozdeľovať energiu do všetkých dostupných foriem rovnomerne (teda aj robiť čo najmenšie rozdiely v preferencii jednej či druhej - kinetickej či potenciálnej - energie. "Presnejšie"teórie ako relativita a kvantová mechanika vec dávajú agendu telesa pri jeho pohybe do súvisu s jeho vlastným časom a zmenou fázy jeho vlnovej funkcie. Naše znalosti prírody ešte nie sú úplné ani tak.

[5]Je k tomu pridaná ešte výhoda ktorú sme tu neilustrovali, len ju spomeňme: Symetrie pomáhajú riešiť diferenciálne rovnice. Ak nájdeme symetrie pre nejaký funkcionál, pomáhajú nám riešiť jeho Euler-Lagrangeove rovnice v istom zmysle dvojnásobne. Teda mať explicitmne formulovaný variačný problém môže byť v istom zmysle výhodnejšie ako mať len rovnice, ktoré ho riešia. (Osobitne ak ani nevieme že za nimi nejaká optimalizácia je.)

Každopádne identifikovať dostatočne všeobecne globálnu agendu systémov je vhodná stratégia k hľadaniu jednotiacich princípov vo fyzike.



Obr. 6: Počiatočné, okrajové a väzbové podmienky sú rozličné, následne aj priebeh pádov vyzerá byť veľmi rôzny. Ale Mesiac padá podľa rovnakej agendy ako jablká. Možno je širšia ako momentálne vidíme, a množstvo ďalších javov je jej zatiaľ nerozpoznaným prejavom.

This project No. 2259/02/01 has received funding from the European Union´s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 945478.