O związkach optymalnego transportu z twierdzeniem o kanapce

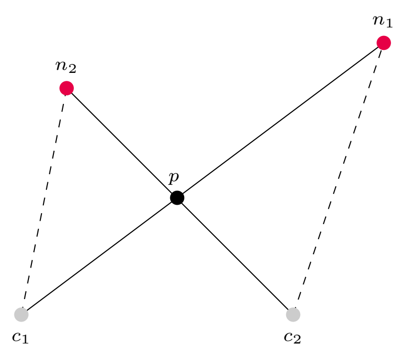
*Krzysztof J. CIOSMAK**

* Uniwersytet Oksfordzki

Badania problemu optymalnego transportu zostały zainicjowane już w 1781 roku przez Gasparda Monge'a, ale na dowód istnienia rozwiązania w przypadku ciągłym trzeba było czekać ponad dwieście lat. Zadanie polega na tym, by znając początkowe i docelowe rozmieszczenie nieodróżnialnych ładunków, tak dobrać docelowe miejsca każdego z nich, by jak najmniej zapłacić. Zakładamy przy tym, że koszt przemieszczenia ładunku z jednego punktu do drugiego jest proporcjonalny do ich odległości.

Łączenie punktów

Zobaczmy, jak powyżej opisany problem jest związany z następującym zagadnieniem. Przypuśćmy, że mamy na płaszczyźnie zadane n punktów szarych i n punktów niebieskich. Chcielibyśmy znaleźć takich n odcinków parami rozłącznych, że każdy z nich ma jeden kraniec szary i jeden kraniec niebieski. Jak łatwo zaobserwować, nie zawsze takie odcinki istnieją. Istotnie, jeśli wszystkie punkty leżą na jednej prostej i wśród nich jest taka trójka jednokolorowych punktów, że pomiędzy nimi nie ma żadnego punktu o odmiennym kolorze, to takich odcinków nie znajdziemy. Aby wyeliminować tego typu sytuacje, będziemy zakładali, że zadane punkty są w *położeniu ogólnym*. Oznacza to, że żadna trójka punktów nie leży na jednej prostej. Pokażemy, że przy tym założeniu można wykazać, że takie odcinki istnieją. Rozważmy bowiem wszystkie możliwe zestawy n odcinków, które łączą punkty szare z niebieskimi w sensie rozważanym powyżej. Takich zestawów jest skończenie wiele – dokładnie: $n!$. Weźmy ten, dla którego suma długości odcinków jest najmniejsza. Przypuśćmy, że w tym wybranym zestawie mamy dwa odcinki, które się przecinają. Sytuacja ta jest zilustrowana na marginesie. Używając oznaczeń z rysunku, widzimy, że nierówność trójkąta implikuje, że



$$d(c_1, n_2) \leqslant d(c_1, p) + d(n_2, p)$$

oraz

$$d(c_2, n_1) \leqslant d(c_2, p) + d(n_1, p).$$

Powyżej $d(x, y)$ oznacza odległość między punktami x i y. Ze względu na poczynione założenie – o tym, że punkty są w położeniu ogólnym – nierówności są ostre. Sumując je, otrzymujemy

$$d(c_1, n_2) + d(c_2, n_1) < d(c_1, n_1) + d(c_2, n_2).$$

Stoi to w sprzeczności z przypuszczeniem, że wybrany zestaw odcinków miał tę własność, że suma długości odcinków była najmniejsza. W istocie, biorąc taki zestaw jak poprzednio, z taką modyfikacją, że zamiast łączyć c_1 z n_1 i c_2 z n_2 łączymy c_1 z n_2 oraz c_2 z n_1 – dostajemy zestaw o ściśle mniejszej sumie długości odcinków. Tym samym wykazaliśmy istnienie pożądanych odcinków.

Rozumowanie, które prowadzi do dowodu tego faktu, jest ściśle związane z opisanym powyżej problemem optymalnego transportu. Rzeczywiście, jeśli pomyślimy o punktach szarych jako o początkowym rozmieszczeniu ładunków, a o punktach niebieskich jako o ich rozmieszczeniu docelowym, to rozważany problem optymalizacyjny jest dokładnie problemem optymalnego transportu.

Lokalizacja

Optymalny transport ma nie tylko bezpośrednie zastosowania. To, co przedstawiliśmy powyżej, stanowi pierwszy krok ku zrozumieniu zastosowania optymalnego transportu do nierówności geometrycznych. Jedną z takich nierówności jest *nierówność izoperymetryczna*. Mówi ona, w klasycznej wersji

Rozwiązanie zadania F 1015.
Na powierzchnię wody w powstałym otworze działa ciśnienie powietrza równe około 10^5 N/m^2, a więc strumień wody będzie wynikiem różnicy ciśnień $\Delta p = 3 \cdot 10^5$ N/m^2. Wyobraźmy sobie, że różnica ciśnień Δp została wytworzona przez pionowy słup wody. Wysokość h takiego słupa wody wynosiłaby

$$h = \frac{\Delta p}{g\rho}$$

i właśnie na taką wysokość wzniósłby się strumień wody wytryskujący z otworu. Po podstawieniu danych otrzymujemy $h \approx 30$ m. Jest to oszacowanie z góry – z powodu oporu powietrza rzeczywista wysokość będzie mniejsza.

znanej już starożytnym, że spośród zbiorów o ustalonej objętości najmniejsze pole powierzchni ma kula.

Taka własność optymalnego transportu, że drogi transportowanych ładunków się nie przecinają (wykorzystaliśmy ją przy rozwiązaniu problemu z odcinkami), pozwala na zredukowanie trójwymiarowego problemu izoperymetrycznego do jego odpowiednika jednowymiarowego. Rozwiązanie tego odpowiednika jest już jednak, jak Czytelnik zechce sprawdzić, oczywiste. Szczegóły metod wykorzystywanych w optymalnym transporcie przy tej redukcji wybiegają jednak poza zakres tego artykułu. Wspomnijmy tylko, że kluczową rolę gra tu *dualność Kantorowicza–Rubinsteina*. Odnotujmy, że Kantorowicz za prace nad tymi zagadnieniami został uhonorowany w 1975 roku Nagrodą Nobla w dziedzinie ekonomii.

Metoda, którą tutaj naszkicowaliśmy, jest zwana lokalizacją. Ma ona szerokie uogólnienie – można ją zastosować także przy badaniu problemu izoperymetrycznego na *rozmaitościach riemannowskich*, które spełniają pewne warunki dotyczące krzywizny i wymiaru. Jest to już jednak materiał na osobną historię.

Problem w trzech wymiarach

Jeden z kierunków, w którym można uogólnić nasze zagadnienie, jest następujący. Załóżmy, że mamy zadane trzy zbiory, każdy po n punktów leżących w przestrzeni trójwymiarowej. Dla wygody nazwijmy punkty pierwszego z tych zbiorów punktami czerwonymi, drugiego – niebieskimi, a trzeciego – zielonymi. Ponownie będziemy zakładali, że zadane punkty są w położeniu ogólnym, a zatem że żadna czwórka punktów nie leży w dwuwymiarowej płaszczyźnie. Naszym zadaniem jest znalezienie n takich trójkątów dwuwymiarowych parami rozłącznych, że każdy z tych trójkątów ma jeden wierzchołek czerwony, jeden niebieski i jeden zielony.

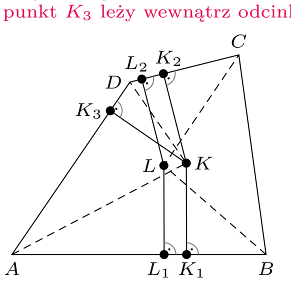
Obecnie nie jest jasne, czy rozwiązanie tego problemu, które byłoby analogiczne do rozwiązania problemu na płaszczyźnie, istnieje. Tym niemniej można udowodnić istnienie pożądanych trójkątów. Metoda, którą się posłużymy, będzie używała znanego twierdzenia Borsuka–Ulama o kanapce. Mówi ono, że gdy mamy kanapkę złożoną z bułki, szynki i sera, to możemy znaleźć taką płaszczyznę, że gdy przetniemy kanapkę wzdłuż niej, to podzielimy każdy ze składników kanapki na dwie równe części.

Idea rozwiązania naszego problemu, pochodząca z pracy Nogi Alona i Jina Akiyamy *Disjoint Simplices and Geometric Hypergraphs*, jest następująca. By móc skorzystać z twierdzenia Borsuka–Ulama, musimy w miejsce zadanych punktów rozważyć małe kulki o środkach w tych punktach. Kulki, których środki są czerwone, będą grały rolę bułki, te o środkach niebieskich – szynki, a zielonych – sera. Twierdzenie nam mówi, że znajdziemy taką płaszczyznę, która każde z tych zbiorów podzieli na dwie równe części. Po drobnej modyfikacji naszego cięcia w razie potrzeby dostaniemy dwa niezależne problemy, identyczne jak problem początkowy, z tym wyjątkiem, że każdy z nich dotyczy zbiorów o liczności nie większej niż $\lceil \frac{n}{2} \rceil$. Tutaj $\lceil x \rceil$ oznacza najmniejszą liczbę całkowitą nie mniejszą od liczby x.

Kontynuując to postępowanie lub – mówiąc bardziej formalnie – stosując zasadę indukcji, uzyskujemy n parami rozłącznych, wypukłych komórek, z których każda zawiera dokładnie po jednym punkcie czerwonym, jednym niebieskim i jednym zielonym. Ze względu na ich wypukłość trójkąty łączące punkty wewnątrz komórek są parami rozłączne. Oczywiście mogą powstać także komórki, które nie zawierają żadnego punktu z zadanych, ale nie stanowią one przeszkody w przeprowadzeniu dowodu istnienia szukanych trójkątów.

Twierdzenie Borsuka–Ulama o kanapce ma swój odpowiednik w przestrzeniach wyżej wymiarowych. Za jego pomocą można też wykazać, że problem dotyczący przestrzeni d-wymiarowej i d różnokolorowych, równolicznych zbiorów, których suma jest w pozycji ogólnej, także ma rozwiązanie.

Rozwiązanie zadania M 1662.
Oznaczmy przez L_1, L_2 rzuty punktu L, odpowiednio, na proste AB i CD, a K_1, K_2, K_3 rzuty punktu K na proste, odpowiednio, AB, CD i DA.
Z wypukłości $ABCD$ punkty K_1 i L_1 leżą wewnątrz odcinka AB, punkty K_2 i L_2 leżą wewnątrz odcinka CD, natomiast punkt K_3 leży wewnątrz odcinka DA.



Mamy zatem równości

$$K_1L_1 = |AB - AK_1 - BL_1|$$

oraz

$$K_2L_2 = |DC - DK_2 - CL_2|.$$

Ponadto

$$AD = AK_3 + DK_3 = AK_1 + DK_2.$$

Analogicznie $BC = BL_1 + CL_2$.
Odcinki K_1L_1 i K_2L_2 są rzutami odcinka KL, odpowiednio, na odcinki AB i CD, skąd $KL \geqslant K_1L_1$ oraz $KL \geqslant K_2L_2$. Wobec tego

$$2KL \geqslant K_1L_1 + K_2L_2 =$$
$$= |AB - AK_1 - BL_1| + |DC - DK_2 - CL_2| \geqslant$$
$$\geqslant |AB - AK_1 - BL_1 + DC - DK_2 - CL_2| =$$
$$= |AB + DC - AD - BC|,$$

gdzie w ostatniej nierówności skorzystaliśmy z nierówności trójkąta.

Uwaga: Niech Ω_K i Ω_L będą okręgami stycznymi do trzech boków czworokąta o środkach, odpowiednio, K i L. Jeśli przez P i Q oznaczymy punkty styczności Ω_K i Ω_L z odcinkiem AB, to z powyższego rachunku można wywnioskować, że

$$PQ = \frac{1}{2}|AB + CD - BC - DA|,$$

skąd teza, gdyż oczywiście $KL \geqslant PQ$.