Spis treści

1 Wstęp		çęp	2	
2	Przejścia fazowe w mechanice statystycznej			
	2.1	Mechanika hamiltonowska	19	
	2.2	Zespół mikrokanoniczny	21	
	2.3	Zespół kanoniczny	23	
	2.4	Jednowymiarowy model Isinga	23	
	2.5	Dwuwymiarowy model Isinga. Argument Peierlsa	28	
3	Per	kolacje i grafy losowe	34	
	3.1	Model Erdösa-Rényi, monotoniczne funkcje boolowskie	34	
	3.2	Spektralna teoria funkcji boolowskich	38	
		3.2.1 Układ Walsha-Fouriera	38	
		3.2.2 Hiperkontrakcja	40	
		3.2.3 Influencje funkcji boolowskich	46	
		3.2.4 Influencje vs. spektrum	47	
		3.2.5 Twierdzenia KKL i Twierdzenia Talagranda	49	
	3.3	Lemat Russo	53	
	3.4	Dowód Twierdzenia Friedguta-Kalaia	55	
4	Per	kolacje na kracie \mathbb{Z}^2	56	
	4.1	Prawdopodobieństwo krytyczne i nierówność FKG	56	
	4.2	Krata dualna	59	
	4.3	Dowód twierdzenia Harrisa	64	
	4.4	Dowód twierdzenia Kestena	68	

Przejścia fazowe

Piotr Nayar *

16-22 września 2013, Małe Ciche

Streszczenie

Opiszemy wybrane aspekty teorii przejść fazowych. Zilustrujemy najważniejsze idee tej teorii i ich uniwersalny charakter. Posłużą nam do tego przykłady naczerpnięte nie tylko z fizyki, ale również z matematyki i informatyki teoretycznej. W szczególności omówimy przejścia fazowe związane z modelem Isinga i modelem perkolacji na kracie całkowitoliczbowej \mathbb{Z}^2 .

1 Wstęp

Przejścia fazowe, rozumiane jako procesy fizyczne, to procesy termodynamiczne polegające na przejściu jednej fazy materii w drugą. Przykładem może być tutaj przejście fazowe ciecz-gaz (parowanie). Tego typu przejścia fazowe możemy opisywać przy pomocy wielkości makroskopowych, bez wnikania w wewnętrzną strukturę materii. Innymi słowy, możemy opisywać parowanie nie wiedząc, że woda składa się z cząsteczek H_2O . Opis ten jest możliwy na gruncie termodynamiki fenomenologicznej, gdzie posługujemy się parametrami termodynamicznymi, takimi jak objętość lub ciśnienie. Czytelnik powinien rozumieć, że ciśnienie jest wielkością, która w zasadzie ma sens jedynie w odniesieniu do eksperymentalnego badania układu złożonego z wielu cząstek. Nie możemy przecież powiedzieć, że gaz złożony np. z dziesięciu cząstek zamkniętych w szczelnym naczyniu ma jakieś ciśnienie. Przez większość czasu cząsteczki te nie mają kontaktu ze ściankami naczynia, a zatem nie oddziałują z nim żadnymi siłami. Nie możemy więc wtedy mówić o dodatnim ciśnieniu. Cząsteczka wywiera ciśnienie na ściankę naczynia tylko w krótkim czasie jej zderzenia z tą ścianką. Jeśli gaz składa się z wielu cząstek, to obserwujemy stałe jego ciśnienie, ponieważ urządzenie pomiarowe nie jest tak dokładne, aby wyłapać i zmierzyć

^{*}Instytut Matematyki, Uniwersytet Warszawski, email: nayar@mimuw.edu.pl

zachodzące szybko w czasie zmiany ciśnienia. W efekcie obserwowane wskazania są wielkościami uśrednionymi po czasie (mówimy, że urządzenie posiada bezwładność).

Wiemy, że w wysokich temperaturach woda przyjmuje formę gazową, a w niskich jest ciałem stałym. Korzystamy z tego wielokrotnie, np. jeżdżąc zimą na nartach. Jeśli mamy określoną ilość wody (np. 1 mol) w stanie równowagi¹, to możemy opisać stan tej ilości materii za pomocą małej liczby parametrów, zwanych parametrami stanu. Co więcej, między parametrami stanu występują związki, zwane równaniami stanu. Np. równanie stanu gazu doskonałego (tzw. równanie Clapeyrona) ma postać

$$pV = NRT$$

gdzie p jest ciśnieniem gazu, T jego temperaturą oraz V – objętością gazu. Stała R jest to tzw. stała gazowa. Oczywiście N oznacza liczbę moli. Stan gazu doskonałego (złożonego z N moli materii) możemy opisać za pomocą parametrów p, V, T, przy czym występuje związek między tymi parametrami. W efekcie, możemy podać tylko dwa z tych parametrów, aby opisać stan, w którym aktualnie znajduje się nasz gaz.

Wróćmy do opisu stanów wody. Przypuśmy, że chcemy opisywać wodę przy pomocy ciśnienia i temperatury, przyjmując, że objętość nie jest kolejnym niezależnym parametrem (podobnie jak to było dla gazu doskonałego). Jeśli zechcemy spróbować uwzględnić rozmiary cząstek i fakt, że cząstki ze sobą oddziałują, to otrzymamy (stosując przybliżenia tzw. pola średniego) równanie stanu gazu van der Waalsa

$$\left(p + N^2 \frac{a}{V^2}\right)(V - Nb) = NRT,$$

ale nawet to równanie nie opisuje wszystkich subtelności gazu rzeczywistego, którego własności zależą między innymi od składu atomowego i symetrii cząstek.

Na Rysunku 1 pokazano diagram fazowy argonu² na płaszczyźnie stanów (T, p). Wyróżnione zostały trzy fazy – stała (s), ciekła (l) i gazowa (g). Gdy ciśnienie wynosi 1 atm, w miarę zwiększania temperatury przechodzimy najpierw przez fazę stałą, potem przez ciekłą, a następnie przez gazową. Mamy zatem dwa przejścia fazowe. Dla ciśnień niższych niż 0,7 atm lód przechodzi od razu w fazę gazową, czyli następuje tzw. sublimacja. Jeśli ciścienie jest wyższe niż 48 atm, nie będziemy potrafili wskazać żadnego konkretnego momentu, w którym podczas zwiększania temperatury ciecz zamieni się w gaz. Jeśli po diagramie fazowym będzie poruszali się po krzywej

¹Tzn. w stanie o ustalonych parametrach takich jak ciścienie, temperatura, objętość, w którym to stanie nie ma makroskopowych przepływów materii, energii...

 $^{^{2}}$ Zajmujemy się tutaj argonem, a nie wodą, ponieważ argon jest jednoatomowym gazem szlachetnym, a zatem ma możliwie prostą strukturę mikroskopową. Okazuje się, że lód wodny ma wiele różnych odmian przy wysokich ciśnieniach.

 γ_1 , zauważymy gwałtowne przejście wody ze stanu ciekłego w gazowy. Natomiast poruszając się po krzywej γ_2 przejście to będzie płynne.



Rysunek 1: Diagram fazowy argonu na płaszczyźnie (T, p).

Na diagramie zaznaczono dwa charakterystyczne punkty – punkt potrójny i punkt krytyczny. W punkcie potrójnym mogą współistnieć ze sobą wszystkie trzy fazy. Punkt krytyczny jest punktem, w którym kończy się linia współistnienia fazy ciekłej i gazowej. Ten drugi punkt będzie nas interesował w sposób szczególny.

Możemy również narysować diagram fazowy wody na płaszczyźnie (v, T), gdzie v = V/N jest objętością molową substancji (Rysunek 2). Na tym diagramie obszary współistnienia faz nie są już liniami. Interesuje nas linia ograniczająca obszar współistnienia fazy ciekłej i gazowej w otoczeniu punktu krytycznego. Okazuje się, że linia ta może być opisana za pomocą zależności potęgowych,

$$v^{(-)} \approx v_0^{(-)} |T - T_c|^{\beta}, \qquad v^{(+)} \approx v_0^{(+)} |T - T_c|^{\beta},$$

przy czym $v^{(-)}$ i $v^{(+)}$ oznaczają odpowiednio lewą i prawą gałąź krzywej. Eksperymentalnie ustalono wartość parametru $\beta = 0.3265$.

Zmieńmy na chwilę temat i zajmijmy się substancjami o własnościach magnetycznych. Tym razem, zamiast ciśnienia, będziemy rozpatrywać pole magnetyczne H, w którym znajduje się układ. Analogiem objętości molowej będzie magnetyzacja materiału³. Tak jak zwiększając ciśnienie gazu w balonie zwiększamy jego objętość, tak

³Czytelnik, który nie wie czym jest magnetyzacja, nie powinien się tym w tej chwili przejmować.



Rysunek 2: Diagram fazowy argonu na płaszczyźnie (v, T).

zwiększając pole magnetyczne zwiększamy magnetyzację materiału. Na tej zasadzie działają kasety VHS. Zapis informacji na takiej kasecie odbywa sie poprzez działanie na taśmę zewnętrznym polem magnetycznym, przez co taśma ulega magnetyzacji. Okazuje się, że (w temperaturach pokojowych) usunięcie pola magnetycznego nie powoduje rozmagnesowania materiału taśmy. Wytworzona magnetyzacja nie znika. Takie materiały nazywamy ferromagnetykami.

Magnetyzacja (czyli wypadkowy moment magnetyczny) jest w ogólności wektorem. Nas będzie interesować model jednoosiowego ferromagnetyka, czyli model, w którym magnetyzacja dowolnego atomu (uwaga: zaczęliśmy mówić o atomach...) jest wektorem równoległym do ustalonej osi. Rysunek 3 pokazuje diagram fazowy jednoosiowego ferromagnetyka na płaszczyźnie (T, H). Jest to analog diagramu (T, p) dla wody. Analogiem fazy ciekłej jest tu faza o dodatniej magnetyzacji, zaś analogiem fazy gazowej jest faza o ujemnej magnetyzacji. Pododnie jak w przypadku wody, idąc krzywą γ_1 zauważymy gwałtowne przejście od fazy (+) do fazy (-). W szczególności

$$\lim_{H \to 0^+} m(T,H) > \lim_{H \to 0^-} m(T,H), \quad T < T_c,$$

gdzie m(T, H) jest magnetyzacją. Natomiast magnetyzacja na krzywej γ_2 jest ciągła.

Rozważmy teraz diagram fazowy na płaszczyźnie (m, T), Rysunek 4. Jest to analog diagramu na płaszczyźnie (v, T) dla wody. Możemy zastanawiać się nad

Zostanie to wyjaśnione później przy opisie mikroskopowym magnetyka.



Rysunek 3: Pogladowy diagram fazowy jednoosiowego ferromagnetyka na płaszczyźnie (T, H).

kształtem linii współistnienia faz (+) oraz (-). Ponownie linia ta w pobliżu punktu krytycznego jest opisywana zależnością potęgową,

$$m(T,0^+) \approx m_0^{(+)} |T - T_c|^{\beta}, \qquad m(T,0^-) \approx m_0^{(-)} |T - T_c|^{\beta}.$$

I tu dzieje się rzecz niewiarygodna. Okazuje się, że wartość eksperymentalna wykładnika β wynosi 0.3265. Wykładnik β jest przykładem wykładnika krytycznego (czyli wykładnika opisującego prawa potęgowe odnoszące się do wielkości fizycznych w otoczeniu punktu krytycznego). Mówiąc trochę nieprecyzyjnie, wykładniki krytyczne mają charakter uniwersalny – nie zależą one od szczegołów układu. Ważny jest jedynie wymiar układu i wymiar tzw. parametru uporządkowania. Temperatura krytyczna T_c oraz amplitudy $m_0^{(+)}$ i $m_0^{(-)}$ nie są uniwersalne. Uniwersalność wykładników krytycznych jest głębokim faktem, który można zrozumieć (przynajmniej z fizycznego, nieścisłego punktu widzenia) za pomocą tzw. grupy renormalizacji.

Zajmiemy się teraz opisem mikroskopowym jednoosiowego ferromagnetyka. Załóżmy, że ferromagnetyk zbudowany jest z N atomów, umiejscowionych w węzłach pewnej sieci lub, ogólniej – pewnego grafu (G, E). Spin (wewnętrzny moment magnetyczny) atomu znajdującego się w wierzchołku v to $s_v \in \mathbb{R}$. Jest to liczba rzeczywista, a nie wektor, ponieważ nasz ferromagnetyk jest jednoosiowy. Teraz musimy skonstruować tzw. hamiltonian. Więcej na temat tego ważnego pojęcia można będzie przeczytać w Podrozdziale 2.1. W tej chwili, jako że jest to poglądowy wstęp, musimy zadowolić się stwierdzeniem, że w naszym przypadku hamiltonian jest energią całkowitą układu, wyrażoną za pomocą spinów. Nasza naczelna zasada jest taka: spiny lubią być skierowane zgodnie z zewnętrznym polem magnetycznym H i lubią być do siebie równoległe. Przyjmijmy zatem, że energia odziaływania spinów s_v i s_w



Rysunek 4: Pogladowy diagram fazowy jednoosiowego ferromagnetyka na płaszczyźnie (m, T).

jest równa $-J_{vw}s_vs_w$, gdzie J_{vw} jest tzw. stałą sprzężenia. Wówczas energetycznie korzystne (układy fizyczne lubią dążyć do minimu energii...) jest, aby spiny s_v i s_w miały ten sam znak (czyli ten sam zwrot). Przyjmijmy, że oddziałują ze sobą wyłącznie spiny znajdujące się w węzłach połączonych krawędzią. Podobnie, niech energia oddziaływania spinu s_v z zewnętrznym polem magnetycznym wynosi $-H_v s_v$, gdzie H_v jest wartością pola w wierzchołku v. Nasz hamiltonian ma zatem postać

$$H = -\sum_{v,w: (v,w)\in E} J_{vw} s_v s_w - \sum_{v\in G} H_v s_v.$$

Należy być świadomym, że w rzeczywistym układzie fizycznym wielkości fizyczne zmieniają się w czasie (np. ciśnienie wywierane przez cząstki na ścianki naczynia). Mamy zatem obserwowaną wartość średnią obserwabli A, ale faktycznie nasza wielkość fizyczna fluktuuje wokół swej wartości średniej. Załóżmy, że chcemy znać wartość średnią pewnej wielkości fizycznej A, zależącej od konfiguracji spinów s_v . Dla przykładu, często bedzie nas interesować średnia magnetyzacja przypadająca na jeden spin,

$$m(\{s_v\}) = \frac{1}{|G|} \sum_{v \in G} s_v.$$

W powyższym wzorze zapis $\{s_v\}$ oznacza pewną konkretną konfigurację spinów, czyli wybór wartości wszystkich zmiennych $s_v, v \in G$. Okazuje się, że obserwowaną wartość średnią można policzyć za pomocą wzoru

$$\langle A \rangle_{kan} = \frac{\int_{G} e^{-\beta H(\{s_v\})} A(\{s_v\}) \, \mathrm{d}\Gamma(\{s_v\})}{\int_{G} e^{-\beta H(\{s_v\})} \, \mathrm{d}\Gamma(\{s_v\})},$$

gdzie Γ jest miarą jednostajną na zbiorze G. Miarę probabilistyczną

$$\mu_H = \frac{1}{Z} e^{-\beta H(\{s_v\})} \mathrm{d}\Gamma,$$

gdzie

$$Z = \int_G e^{-\beta H(\{s_v\})} \,\mathrm{d}\Gamma(\{s_v\})$$

jest tzw. sumą statystyczną, nazywamy miarą Gibbsa. Innymi słowy (to zdanie jest kierowane do tych Czytelników, którzy jeszcze nie spotkali się z pojęciem miary), wartość średnia obserwabli A jest ważoną średnią arytmetyczną wartości $A(\{s_v\})$ z wagami $e^{-\beta H(\{s_v\})}$.

Ale zaraz, moment! Sytuacja fizyczna jest taka, że spiny zmieniają swoje wartości w czasie, a zatem wartość obserwabli A podlega fluktuacjom. Rzecz się ma dokładnie tak samo, jak w przypadku ciśnienia gazu zamkniętego w naczyniu – z mikroskopowego punktu widzenia ciśnienie jest pewną bardzo skomplikowaną funkcją czasu, a nas interesuje tylko wartość średnia tej wielkości, którą mierzy urządzenie pomiarowe. Ale jest to średnia po czasie, a my przecież definiując miarę Gibbsa i podając regułę liczenia średniej wartości A nic o czasie nie mówiliśmy! Wszystko dlatego, że podstawą mechaniki statystycznej jest przyjęcie następującego założenia, zwanego hipotezą ergodyczną,

> Dla układu znajdującego się w temperaturze T wartość średnia obserwabli liczona względem czasu jest dla dużych czasów równa wartości średniej tej obserwabli względem miary Gibbsa.

Prawdziwość tego założenia można zweryfikować w sposób ścisły jedynie dla kilku bardo prostych układów. Więcej informacji na temat opisanego sposobu obliczania średnich wartości obserwabli znajdzie Czytelnik w Podrozdziałach 2.2 i 2.3.

Interesować nas będzie tzw. model Isinga. W modelu tym graf (G, E) jest podzbiorem kraty całkowitoliczbowej \mathbb{Z}^d , przy czym wierzchołki v, w są połączone krawędzią wtedy i tylko wtedy gdy |w - v| = 1, gdzie $|\cdot|$ jest odległością euklidesową. Mówimy czasem, że oddziałują ze sobą jedynie spiny będące najbliższymi sąsiadami



Rysunek 5: Krata \mathbb{Z}^2 .

(Rysunek 5). Przyjmujemy również, że spiny mogą przyjmować wyłącznie wartości ze zbioru dwuelementowego $\{-1, 1\}$. Będziemy dalej przyjmować, że $V = \{1, \ldots, N\}^d$. W Rozdziale 2.4 rozważymy układ jednowymiarowy (d = 1). Rozważymy przypadek periodycznych warunków brzegowych (czyli przypadek N spinów na dyskretnym okręgu) i dodatnich warunków brzegowych. Zbadamy te układy wyznaczając sumy statystyczne.

Ważne będzie dla nas pojęcie tzw. granicy termodynamicznej. Procedura przejścia do granicy termodynamicznej polega na rozważeniu układu o skończonych rozmiarach, obliczeniu interesującej nas wielkości fizycznej, np. średniej magnetyzacji przypadającej na jeden spin, a następnie na przejściu z rozmiarem układu do nieskończoności. Chodzi nam o to, aby wydobyć człony związane z zachowaniem się układu w jego wnętrzu, a nie przy ściankach ograniczających układ. W przypadku modelu Isinga przejście do granicy termodynamicznej bedzie oznaczało wzięcie granicy $N \to \infty$. Niech m(N, H, T) bedzie średnią magnetyzacją obliczoną za pomocą miary Gibbsa dla skonczonego układu. Definiujemy

$$m(H,T) = \lim_{N \to \infty} m(N,H,T).$$

Funkcja m(H,T) opisuje średnią magnetyzację w układzie jednorodnym. Możemy teraz powrócić do Rysunku 4. Równania krzywej ograniczającej obszar współistnienia faz (+) oraz (-) mają postać

$$m(T, 0^+) = \lim_{H \to 0^+} m(H, T), \qquad m(T, 0^-) = \lim_{H \to 0^-} m(H, T).$$

Dla $T > T_c$ mamy $m(T, 0^+) = m(T, 0^-) = 0$. Posiłkując się przykładem kasety VHS możemy stwierdzić, że w dużych temperaturach na takiej kasecie nie może być nic nagrane. Niezerowa wartość m(H, T) powstaje w momencie zapisu, przy pomocy oddziaływania z zewnętrznym polem magnetycznym. Jeśli jednak czynnik ten zostanie usunięty, magnetyzacje powróci do wartości zerowej. Odpowiada to braniu granic $m(T, 0^{\pm})$. Jeśli jednak $T < T_c$, mamy $m(T, 0^{+}) > 0 > m(T, 0^{-})$, czyli działając np. dodatnim polem magnetyczny, a następnie usuwając to pole, otrzymamy układ z dodatnią magnetyzacją. Temperatura T_c jest często nazywana temperaturą Curie.

W Podrozdziale 2.4 udowodnimy, że dla modelu jednowymiarowego obszar ten jest zdegenerowany, czyli $m(T, 0^+) = m(T, 0^-) = 0$ dla T > 0. Zatem w tym przypadku układ nie przejawia własności ferromagnetycznych i nie jest możliwe współistnienie fazy (+) i fazy (-). Pokażemy również, że magnetyzacja w układzie z zerowym polem i z dodatnimi warunkami brzegowymi znika w granicy termodynamicznej. Policzymy gęstość tzw. energii swobodnej układu w temperaturze T i polu magnetycznym H przypadającej na jeden spin,

$$f(H,T) = -k_B T \lim_{N \to \infty} \frac{\ln Z_N(T,H)}{N}$$

i wykażemy, że jest to analityczna funkcja parametrów T, H. W myśl ogólnej teorii przejść fazowych oznacza to brak punktu krytycznego dla T > 0.

W Podrozdziale 2.5 zajmiemy się dwuwymiarowym modelem Isinga z dodatnimi warunkami brzegowymi i z zerowym polem magnetycznym. Zaprezentujemy tzw. argument Peierlsa i wykażemy, że magnetyzacja w granicy termodynamicznej jest dodatnia. W przypadku swobodnych warunków brzegowych pokażemy, że gęstość energia swobodnej w granicy termodynamicznej nie jest różniczkowalną funkcją zmiennej H. Oznacza to obecność linii współistnienia fazy (+) o ściśle dodatniej magnetyzacji i fazy (-) o ściśle ujemnej magnetyzacji.

Warto tu dodać, że dwuwymiarowy model Isinga dla H = 0 może być ściśle zbadany, tzn. możliwe jest wyznaczenie gęstości energii swobodnej f = f(T, 0) dla tego układu,

$$-\frac{f}{k_B T} = \ln 2 + \frac{1}{2\pi^2} \int_0^{\pi} \int_0^{\pi} \ln\left((\cosh(2K)^2 - \sinh(2K)(\cos x + \cos y))\right) \,\mathrm{d}x \,\mathrm{d}y,$$

gdzie $K = \frac{J}{k_B T}$. Zauważmy, że

$$\cosh(2K)^{2} - \sinh(2K)(\cos x + \cos y) \ge \cosh(2K)^{2} - 2\sinh(2K)$$
$$= 1 + \sinh^{2}(2K) - 2\sinh(2K)$$
$$= (1 - \sinh(2K))^{2} \ge 0.$$

Widzimy, że jeśli dla temperatury T mamy $\sinh(2K) \neq 1$, bo funkcja f jest analityczna w otoczeniu tego punktu. Istnieje zatem dokładnie jedna wartość K, czyli dokładnie jedna wartość T, dla której funkcja f posiada osobliwość⁴. Rozwiązując równanie $\sinh(2K) = 1$ otrzymujemy

$$T_c = \frac{2J}{k_B} \cdot \frac{1}{\ln(1+\sqrt{2})}.$$

Możliwe jest również wyznaczenie średniej magnetyzacji przypadającej na jeden spin,

$$m(T,0^{+}) = -m(T,0^{-}) = \begin{cases} \left(1 - \sinh^{-4}\left(\frac{2J}{k_{B}T}\right)\right)^{1/8} & \text{gdy} \quad T < T_{c} \\ 0 & \text{gdy} \quad T \ge T_{c} \end{cases}$$

Zauważmy, że wzór dla $T < T_c$ możemy zapisać w formie

$$m(T, 0^+) = \left(1 - \sinh^{-4}\left(\frac{T}{T_c}\ln(1 + \sqrt{2})\right)\right)^{1/8}.$$

Na Rysunku 6 pokazano wykresy $m(T, 0^+)$ i $m(T, 0^-)$ w funkcji T/T_c . Krzywe ograniczają obszar współistnienia faz (+) oraz (-). Otrzymaliśmy wykres zgodny z poglądowym Rysunkiem 4.



Rysunek 6: Wykresy $m(T, 0^+)$ i $m(T, 0^-)$ w funkcji T/T_c .

Możemy również wyznaczyć z łatwością wykładnik krytyczny β , czyli wykładnik opisujący zachowanie się funkcji $m(T, 0^+)$ w otoczeniu $T = T_c$. Okazuje się, że

$$m(T, 0^+) \propto |T - T_c|^{1/8}$$

⁴Trzeba chwilę popracować, żeby stwierdzić, że faktycznie nie mamy analityczności dla tej krytycznej wartości temperatury.

czyli $\beta = 1/8$. Dlaczego nie otrzymaliśmy $\beta = 0.3265$? Bo rozważyliśmy układ dwuwymiarowy, a podane wartości eksperymentalne odnosiły się do układów trójwymiarowych – wspomnieliśmy już, że wykładnik β zależy od wymiaru układu.

Do tej pory, z matematycznego punktu widzenia, badaliśmy własności miary Gibbsa na zbiorze $\{-1,1\}^n$, gdzie M było liczbą spinów w układzie. W drugiej części wykładu, Rozdział 3 zajmiemy się badaniem miar produktowych na kostce $\{-1, 1\}^n$ i funkcji $f: \{-1,1\}^n \to \mathbb{R}$. Naszym pierwszym celem będzie udowodnienie przybliżonego prawa 0-1 Kołmogorowa dla modelu grafu losowego Erdösa-Renyi. W modelu tym rozważamy zbiór złożony z N wierzchołków. Każda para wierzchołków jest połączona krawędzią z prawdopodobieństwem $p \in [0, 1]$, niezależnie. Niech x_i będzie zmienną przyjmującą wartość 1, jeśli *i*-ta krawędź występuje w grafie oraz -1, jeśli *i*-tej krawędzi w grafie nie ma. Zdefiniowaliśmy graf losowy, którego rozkład krawędzi jest miarą $\mathbb{P}_p = (p\delta_1 + (1-p)\delta_1)^{\otimes n}$, gdzie $n = \binom{N}{2}$. Interesują nas symetryczne zdarzenia monotoniczne związane z tym grafem. Zdarzenie jest monotoniczne, jeśli dołożenie krawędzi nie może spowodować, że zdarzenie przestanie zachodzić. Zdarzenie A jest symetryczne, jeśli jest ono niezmiennicze ze względu na przenumerowanie wierzchołków. Przykładem symetrycznych zdarzeń monotonicznych są zdarzenia graf jest spójny oraz w grafie istnieje klika K_5 . Zauważmy, że nasz model zależy od parametrów p i n. Interesuje nas zachowanie funkcji $[0,1] \ni p \mapsto \mathbb{P}_p(A)$. Oczywiście spodziewamy się, że funkcja ta jest ciągła i monotoniczna dla zdarzeń monotonicznych. Niech $\varepsilon \in (0, 1/2)$. Definiujemy $p_{-}(\varepsilon)$ i $p_{+}(\varepsilon)$ za pomocą relacji $\mathbb{P}_{p_{-}(\varepsilon)} = \varepsilon$ oraz $\mathbb{P}_{p_+(\varepsilon)} = 1 - \varepsilon$. Okazuje się, że przedział $(p_-(\varepsilon), p_+(\varepsilon))$ jest dla dużych n bardzo krótki, mianowicie

$$|p_+(\varepsilon) - p_-(\varepsilon)| \le c_1 \frac{\ln(1/2\varepsilon)}{\ln n}.$$

Jest to właśnie przybliżone prawo 0-1 Kołmogorowa. Możemy zatem powiedzieć, że monotoniczne zdarzenia symetryczne podlegają przejściom fazowym. Mamy zasadniczo dwie fazy – fazę 0, w której zdarzenie A nie zachodzi i fazę 1, w której zdarzenie zachodzi. Dla małych prawdopodobieństw p wystąpienie zdarzenia A jest mało prawdopodobne, zaś dla dużych p jest bliskie 1, przy czym funkcja $[0,1] \ni p \mapsto \mathbb{P}_p(A)$ rośnie gwałtownie na bardzo krótkim przedziale. Fakt ten zostanie udowodniony w Podrozdziale 3.4 (Twierdzenie 2). Dowód zostanie sformułowany w języku funkcji boolowskich $f : \{-1,1\}^n \to \{-1,1\}$. Będzie on od nas wymagał wypracowania technik związanych z analizą spektralną funkcji na kostce dyskretnej. Każdą funkcję $f : \{-1,1\}^n \to \mathbb{R}$ możemy zapisać jednoznacznie w postaci wielomianu

$$f = \sum_{S \subset \{1, \dots, n\}} a_S w_S,$$

gdzie $w_S(x_1, \ldots, x_n) = \prod_{i \in S} x_i$. Współczynniki $(a_S)_{S \subset S}$ nazywamy spektrum funkcji f. Kluczowy będzie dla nas operator szumu,

$$T_{\delta}(f) = \sum_{S \subset \{1, \dots, n\}} a_S w_S \delta^{|S|}.$$

Wykażemy, że operator ten jest hiperkontraktywny, czy spełnia nierówność

$$||T_{\delta}f||_2 \le ||f||_{1+\delta^2},$$

gdzie $||f||_p = (\mathbb{E}|f|^p)^{1/p}$. Ostatnim krokiem pośrednim w dowodzie Twierdzenia 2 będzie Twierdzenie KKL. W sformułowaniu tego twierdzenia występują wielkości

$$I_i^p(f) = \mathbb{P}_p\left(f(x) \neq f(x^i)\right),\,$$

gdzie $x^i = (x_1, \ldots, -x_i, \ldots, x_n)$ dla $x = (x_1, \ldots, x_n)$. Wielkości I_i^p , $i = 1, \ldots, n$ nazywamy influencjami funkcji f. Opisują one średni wpływ *i*-tego bitu na wartość funkcji f. Pokażemy nierówność Talagranda

$$\sum_{i=1}^{n} I_i^p(f) \ge c \operatorname{Var}_p(f) \ln\left(\frac{1}{\max_{i=1,\dots,n} I_i^p(f)}\right).$$

Nierówność ta, w połączeniu z tzw. Lematem Russo, w łatwy sposób implikuje Twierdzenie 2. Zgodnie z Lematem Russo dla funkcji monotonicznych zmiany funkcji $[0,1] \ni p \mapsto \mathbb{P}_p(A)$ możemy kontrolować przez sumę influencji funkcji f, a konkretnie

$$\frac{\mathrm{d}}{\mathrm{d}p}\mathbb{P}_p(f=1) = \sum_{i=1}^n I_i^p(f).$$

W Rozdziale 4 będziemy badać pewien graf losowy o nieskończonej liczbie wierzchołków. Rozważmy kratę \mathbb{Z}^2 ze strukturą grafu, w której krawędziami połączone są sąsiednie wierzchołki (każdy punkt ma 4 sąsiednie wierzchołki – rzecz się ma podobnie jak w modelu Isinga). Następnie, niezależnie, o każdej krawędzi decydujemy, czy jest ona otwarta (z prawdopodobieństwem p), czy też zamknięta (z prawdopodobieństwem 1 - p). Otrzymaliśmy model podobny do modelu Erdösa-Renyi, z tym że przestrzeń krawędzi jest teraz nieskończona i nie każdej parze wierzchołków odpowiada krawędź. Niech E_{∞} będzie zdarzeniem polegającym na istnieniu nieskończonej otwartej spójnej składowej. Otwarta spójna składowa (czyli otwarty *klaster*) jest zbiorem wierzchołków V o tej własności, że dla dowolnych $a, b \in V$ wierzchołki a, b da się połączyć ścieżką złożoną z otwartych krawędzi. Z klasycznego prawa 0-1 Kołmogorowa wynik, że $\mathbb{P}_p(E_{\infty}) \in \{0,1\}$. Okazuje się, że dla $p \leq 1/2$ mamy



Rysunek 7: Pogladowy wykres funkcji $p \mapsto \theta(p)$.

 $\mathbb{P}_p(E_{\infty}) = 0$ oraz dla p > 1/2 jest $\mathbb{P}_p(E_{\infty}) = 1$. Jest to słynne twierdzenie Harrisa-Kestena. Naszym celem jest podanie dowodu tego twierdzenia.

Posłużymy się parametrem $\theta(p) = \mathbb{P}_p(|C_0| = \infty)$, gdzie C_0 jest otwartą spójną składową punktu 0. Na Rysunku 7 pokazano poglądowy wykres funkcji $p \mapsto \theta(p)$. Wielkość $\theta(p)$ możemy traktować jak parametr uporządkowania, który rozróżnia dwie fazy – fazę bez nieskonczonego otwartego klastra ($p \leq 1/2$) i fazę z nieskończonym otwartym klastrem (p > 1/2). Analogiem obserwabli

$$m(\{s_v\}) = \frac{1}{|G|} \sum_{v \in G} s_v$$

w modelu Isinga jest tutaj wielkość $\mathbf{1}_{\{|C_0|=\infty\}}$. Podobnie jak $m(\{s_v\})$ zależy od konkretnej konfiguracji spinów, tak $\mathbf{1}_{\{|C_0|=\infty\}}$ zależy od wyboru otwartych krawędzi. Natomiast $m = \langle m(\{s_v\}) \rangle_{kan}$ i $\theta(p) = \mathbb{E} \mathbf{1}_{\{|C_0|=\infty\}}$ są wielkościami uśrednionymi rozróżniającymi fazy. Parametr p może być porównany do parametru T w modelu Isinga. Punkt p_c nazywamy prawdopodobieństwem krytycznym – w modelu Isinga odpowiada mu temperatura krytyczna T_c . Zachowanie się funkcji $p \mapsto \theta(p)$ można opisać zależnością potęgową

$$\theta(p) \approx (p - p_c)^{\frac{5}{36}}.$$

Zatem $\beta = 5/36$ jest odpowiednim wykładnikiem krytycznym. Okazuje się, że jest on uniwersalny, tzn. nie zależy od szczegółów sieci. Oznacza to, że wykładnik β bedzie taki sam dla kraty heksagonalnej, kraty trójkątnej i wielu innych dwuwymiarowych krat na \mathbb{R}^2 .

Zwróćmy uwagę na ważne pojęcie funkcji korelacji dla modelu Isinga,

$$G(x,y) = \langle s_x s_y \rangle_{kan} - \langle s_x \rangle_{kan} \langle s_y \rangle_{kan} \,.$$

nazwa	definicja	zachowanie	wyk.	wartość na \mathbb{Z}^2
funkcja per- kolacji	$\theta(p) = \mathbb{P}_p\left(C_0 = \infty\right)$	$\theta(p) \approx (p - p_c)^{\beta}$	β	5/36
średnia wiel- kość C_0	$\chi^f(p) = \mathbb{E}_p(C_0 ; C_0 < \infty)$	$\chi^f(p) \approx p - p_c ^{-\gamma}$	γ	43/18
liczba otwartych klastrów na wierzchołek	$\kappa(p) = \mathbb{E}_p(C_0 ^{-1})$	$\kappa'''(p) \approx p - p_c ^{-1-\alpha}$	α	-2/3
$\begin{array}{c} \text{momenty} \\ C_0 \end{array}$	$\chi_k^f(p) = \mathbb{E}_p(C_0 ^k; C_0 < \infty)$	$\frac{\chi_{k+1}^f(p)}{\chi_k^f(p)} \approx p - p_c ^{-\Delta}$	Δ	91/36
długość kore- lacji	$\xi(p)^{-1} = \frac{-1}{n} \lim_{n \to \infty} \ln \tau_p^f(ne_1)$	$\xi(p) \approx p - p_c ^{-\nu}$	ν	4/3
rozkład $ C_0 $	$\mathbb{P}_{p_c}(C_0 =n)$	$ \mathbb{P}_{p_c}(C_0 = n) \approx $ $n^{-1-1/\delta} $	δ	91/5
średnica $ C_0 $	$rad(C_0) = \max\{ x : x \in C_0\}$	$\mathbb{P}_{p_c}(rad(C_0) = n) \approx n^{-1-1/\rho}$	ρ	96/10
funkcja spój- ności	$\tau_p^f(x) = \mathbb{P}_p\left(x \in C_0; C_0 < \infty\right)$	$\tau^f_{p_c}(ne_1) \approx n^{2-d-\eta}$	η	5/24

Tablica 1: Prawa potęgowe i odpowiadające im wykładniki krytyczne wraz z wartościami liczbowymi dla perkolacji na kracie \mathbb{Z}^2 .

nazwa	${ m definic}$ ja	zachowanie	wyk.	wartość na \mathbb{Z}^2
spontaniczna magnetyza- cja	$m(T,H) = -\left(\frac{\partial f}{\partial H}\right)_T$	$m(T,0^+) \approx T-T_c ^{\beta}$	β	1/8
podatność magnetyczna	$\chi(H,T) = \left(\frac{\partial m}{\partial H}\right)_T$	$\chi(0,T) \approx T - T_c ^{-\gamma}$	γ	7/4
ciepło właściwe	$C_H = T^2 \left(\frac{\partial^2 g}{\partial T^2}\right)_H$	$C_{H=0} \approx T - T_c ^{-\alpha}$	α	0
$ \begin{array}{c} \text{magnetyzacja} \\ \text{w} \ T = T_c \end{array} $	$m(T,H) = -\left(\frac{\partial f}{\partial H}\right)_T$	$m(T_c, H) \approx H ^{1/\delta}$	δ	15
	$g_s(\lambda^a \tau, \lambda^b h) = \lambda g_s(\tau, h)$	$\Delta = b/a$	Δ	15/8
długość kore- lacji	$g(x,y)_{H=0} \approx_{T \to T_c} \frac{e^{- x-y /\xi}}{ x-y ^{d-2+\eta}}$	$ \begin{array}{ll} \xi(T > T_c, H = 0) \\ T - T_c ^{-\nu} \end{array} \end{array} \approx $	ν	1
funkcja kore- lacji	$g(x,y)_{H=0} = Cor(s_x, s_y)$	$g(x,y) \approx_{T \to T_c} \frac{e^{- x-y /\xi}}{ x-y ^{d-2+\eta}}$	η	1/4

Tablica 2: Prawa potęgowe i odpowiadające im wykładniki krytyczne wraz z wartościami liczbowymi dla modelu Isinga na kracie \mathbb{Z}^2 .

Okazuje się, że funkcja ta ma następujące asymptotyczne zachowanie dla H = 0,

$$g(x,y)_{H=0} \approx_{T \to T_c} \frac{e^{-|x-y|/\xi}}{|x-y|^{d-2+\eta}}$$

Parametr ξ jest tzw. długością korelacji. Jego zachowanie w otoczeniu punktu krytycznego może być opisane zależnością potęgową,

$$\xi(T, H=0) \approx |T - T_c|^{-\nu}.$$

W przypadku modelu dwuwymiarowe mamy $\nu = 1$. Zauważmy, że dla $T > T_c$ korelacja spinów maleje wykładniczo wraz z ich odległością. Gdy jednak znajdziemy się w punkcie krytycznym, otrzymujemy $\xi(T_c) = +\infty$ i funkcja korelacji zanika potęgowo,

$$g(x,y)_{H=0,T=T_c} \approx \frac{1}{|x-y|^{d-2+\eta}}.$$

Funkcję ξ możemy interpretować jako charakterystyczną skalę długości, występujacą w układzie. Spiny, których wzajemna odległość jest rzędu ξ mają korelację, która jeszcze nie jest eksponencjalnie mała względem |x - y|. Ten fakt leży u podstaw stosowalności metody renormalizacji, za pomocą której można wyprowadzić

relacje skalowania	relacje hiperskalowania
$\gamma = \nu(2 - \eta)$	$d\rho = \delta + 1$
$2 - \alpha = \gamma + 2\beta = \beta(\delta + 1)$	$d\nu = 2 - \alpha$
$\Delta = \delta\beta$	

Tablica 3: Relacje skalowania i hiperskalowania w modelu perkolacji i modelu Isinga dla wymiaru d.

tzw. relacje hiperskalowania. Natomiast relacje skalowania niezależne od wymiaru wynikają z hipotezy skalowania części osobliwej energii swobodnej Gibbsa. Według hipotezy skalowania, istnieje rozkład energii swobodnej Gibbsa na część regularną i część singularną,

$$g(\tau, H) = g_{reg}(\tau, H) + g_s(\tau, H), \qquad \tau = \frac{T - T_c}{T_c},$$

czy czym część singularna ma charakter uniwersalny i podlega skalowaniu

$$g_s(\lambda^a \tau, \lambda^b H) = \lambda g_s(\tau, H)$$

Zdefiniujmy $\Delta = b/a$. Można wówczas pokazać (zainteresowanemu Czytelnikowi polecamy to zadanie jako ćwiczenie), licząc wielkości podane w Tabeli 2, że prawdziwe są relacje

$$\begin{array}{rcl} \alpha &=& 2-1/a \\ \beta &=& 1/a - \Delta \\ \gamma &=& 2\eta - 1/a \\ \delta &=& \Delta/\beta \end{array}$$

Można stąd otrzymać trzy spośród czterech relacji skalowania podanych w Tabeli 3. Analogiczna wielkość ξ dla modelu perkolacji opisuje zachowanie się funkcji

$$\tau_p^f(x) = \mathbb{P}_p(x \in C_0; \ |C_0| < \infty),$$

gdzie C_0 jest otwartą spojną składową punktu 0. Niech $e_1 = (1, 0, ..., 0)$. Długość korelacji definiujemy jako

$$\xi(p)^{-1} = \frac{-1}{n} \lim_{n \to \infty} \ln \tau_p^f(ne_1).$$

Oznacza to, że $\tau_p^f(ne_1)$ maleje wykładniczy, przy czym charakterystyczna skala długości jest zadana właśnie funkcją ξ . Dla $p \to p_c^-$ mamy

$$\xi(p) \approx |p - p_c|^{-\nu}$$

i ponownie długość korelacji dąży do $+\infty$, gdy zbliżamy się do punktu krytycznego.

Bardzo wiele ważnych tematów nie zostało poruszonych w tym wstępie, nad czym autor bardzo ubolewa. W szczególności, nie została przedstawiona teoria renormalizacji (która nie jest teorią ścisłą z matematycznego punktu widzenia) i nie zostały opisane twierdzenia związane z konforemną niezmienniczością perkolacji i modelu Isinga. Nie zostały wprowadzone procesy SLE, opisujące krzywe ograniczające klastry w granicy $\delta \rightarrow 0$, gdzie δ jest stałą sieci. Wreszcie, nie została wyjaśniona przyczyna istnienia głębokich związków między modelem Isinga i perkolacjami. Nie ma na te rzeczy miejsca w niniejszym wykładzie.

2 Przejścia fazowe w mechanice statystycznej

2.1 Mechanika hamiltonowska

Rozważmy zespół cząstek podlegających prawom mechaniki klasycznej. W chwili czasowej t układ ten może być opisany poprzez podanie uogólnionych położeń cząstek⁵ $x_1, \ldots, x_N \in \mathbb{R}$ oraz ich uogólnionych pędów p_1, \ldots, p_N . Oznacza to, że zachowanie układu w czasie opisane jest przez równania Hamiltona

$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = -\frac{\partial H}{\partial q_i} \\ \frac{\mathrm{d}q_i}{\mathrm{d}t} = \frac{\partial H}{\partial p_i}.$$

gdzie H jest tzw. hamiltonianem. Rozwiązanie tego układu 2N równań opisuje ewolucję czasową w przestrzeni fazowej $\Gamma \subset \mathbb{R}^{2N}$.

W przypadku n cząstek swobodnych poruszających się w przestrzeni eulidesowej \mathbb{R}^3 mamy $X_i = (x_i^{(1)}, x_i^{(2)}, x_i^{(3)}) \in \mathbb{R}^3$ oraz $P_i = (p_i^{(1)}, p_i^{(2)}, p_i^{(3)}) \in \mathbb{R}^3, i = 1, \ldots, n.$ W tej sytuacji N = 3n. Mamy wówczas 3N położeń $x_i^{(j)}, i = 1, \ldots, n, j = 1, 2, 3$ oraz 3N pędów $p_i^{(j)}, i = 1, \ldots, n, j = 1, 2, 3$. W danej chwili czasowej położenie układu opisuje punkt w przestrzeni fazowej $\Gamma = \mathbb{R}^{6n}$. W przypadku braku odziaływać układu z otoczeniem, hamiltonian H jest energią układu wyrażoną przy pomocy położeń i pędów.

Przykład 1. Rozważmy swobodną cząstkę poruszającą się po prostej rzeczywistej \mathbb{R} . W tej sytuacji mamy jedno położenie uogólnione $x \in \mathbb{R}$ oraz jeden pęd uogólniony $p \in \mathbb{R}$. Przetrzenią fazową jest zatem $\Gamma = \mathbb{R}^2$. Energia cząstki składa się wyłącznie z energii kinetycznej $\frac{1}{2}mv^2$, gdzie v jest prędkością cząstki. Pamiętamy, że p = mv, a zatem hamiltonian wyrażony przez pęd i położenie jest równy

$$H(x,p) = \frac{p^2}{2m},$$

a zatem nie zależy od x. Równania Hamiltona mają postać

$$\frac{\frac{\mathrm{d}p}{\mathrm{d}t} = 0}{\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{p}{m}}, \; ,$$

Wynika stąd, że $p(t) = p_0$ i $x(t) = \frac{p_0}{m}t + x_0$. Zauważmy, że $v_0 = p_0/m$ jest prędkością naszej cząstki swobodnej. Otrzymaliśmy ruch jednostajny.

⁵Jeśli cząstka porusza się po okręgu, położeniem uogólnionym może być wtedy zmienna kątowa $\theta \in [0, 2\pi)$ opisująca pozycję cząstki na tym okręgu.

Przykład 2. Załóżmy, że nasza cząstka nie jest swobodna, ale porusza się w pewnym potencjale V(x). Całkowita energia cząstki jest teraz równa

$$H(x,p) = \frac{p^2}{2m} + V(x).$$

Równania Hamiltona mają postać

$$\frac{\frac{\mathrm{d}p}{\mathrm{d}t}}{\frac{\mathrm{d}x}{\mathrm{d}t}} = -V'(x) ,$$
$$\frac{\frac{\mathrm{d}x}{\mathrm{d}t}}{\frac{\mathrm{d}x}{\mathrm{d}t}} = \frac{p}{m}. ,$$

Podstawiając drugie równanie do pierwszego otrzymujemy

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -V'(x).$$

W powyższym równaniu $\frac{d^2x}{dt^2}$ ma interpretację przyspieszenia a zaś -V'(x) jest siłą działającą na naszą cząstkę. Otrzymaliśmy zatem drugą zasadę dynamiki Newtona, F = ma.

Podamy ogólną metodę rozwiązywania tego równania. Niech $v(t) = \frac{dx}{dt}(t)$. Wówczas nasze równanie jest równoważne z

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{mv^2}{2} + V(x)\right) = 0.$$

Zatem wielkość

$$E = \frac{mv^2}{2} + V(x)$$

jest zachowana podczas ruchu. Zauważmy, że E jest całkowitą energią układu. Otrzymaliśmy zachowanie hamiltonianu w czasie. Ogólnie, przy braku sił zewnętrznych, hamiltonian jest zawsze wielkością zachowaną. Brak sił zewnętrznych oznacza brak jawnej zależności hamiltonianu od czasu. Możemy bardzo łatwo udowodnić powyższy fakt korzystając z reguły łańcuchowej,

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \sum_{i=1}^{N} \frac{\partial H}{\partial p_i} \frac{\mathrm{d}p_i}{\mathrm{d}t} + \frac{\partial H}{\partial q_i} \frac{\mathrm{d}q_i}{\mathrm{d}t} = \sum_{i=1}^{N} -\frac{\partial H}{\partial p_i} \frac{\partial H}{\partial q_i} + \frac{\partial H}{\partial q_i} \frac{\partial H}{\partial p_i} = 0.$$

Skorzystaliśmy z równań Hamiltona.

Biorąc pod uwagę fakt, że $v = \frac{dx}{dt}$, otrzymujemy

$$\left|\frac{\mathrm{d}x}{\mathrm{d}t}\right| = \sqrt{\frac{2}{m}}(E - V(x)).$$

Jest to w zasadzie równanie różniczkowe 1-rzędu o rozdzielonych zmiennych.

Przykład 3. Rozważmy Przykład 2 z potencjałem $V(x) = \frac{1}{2}kx^2$. Jest to tzw. potencjał oscylatora harmonicznego. Otrzymujemy równanie różniczkowe liniowe rzędu 2,

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + kx = 0$$

Czytelnik zaznajomiony z teorią równań rożniczkowych zwyczajnych bez trudu znajdzie rozwiązanie tego równania,

$$x(t) = A\sin(\omega t + \alpha),$$

gdzie $\omega=\sqrt{k/m},$ zaśAi α są pewnymi stałymi, zależącymi od położenia i prędkości początkowej.

Przykład 4. Ostatni przykład dotyczy cząstki poruszającej się w polu grawitacyjnym. W tym przypadku cząstka opisywana jest przez współrzędne x, y, z oraz pędy p_x, p_y, p_z . Hamiltonian ma postać

$$H(x, y, z, p_x, p_y, p_z) = \frac{p_x^2}{2m} + \frac{p_y^2}{2m} + \frac{p_z^2}{2m} + mgz.$$

Czytelnik zechce samodzielnie napisać i rozwiązać równania Hamiltona oraz przekonać się, że otrzymujemy znane ze szkoły równania rzutu ukośnego (należy jednak pamiętać, że w szkole zwykle rzut ukośny rozpatrywany jest jedynie w płaszczyźnie rzutu i dlatego mamy wtedy tylko dwa wymiary).

2.2 Zespół mikrokanoniczny

Rozważmy teraz podzbiór A przestrzeni fazowej $\Gamma \subset \mathbb{R}^m$. Równania Hamiltona można zapisać ogólnie w postaci $\frac{dx}{dt} = A(x)$, gdzie $A : \mathbb{R}^m \to \mathbb{R}^m$ jest pewnym polem wektorowym (w naszym przypadku pole to jest generowane przez zadany hamiltonian). Równanie to generuje potok fazowy (ϕ_t)_{t ≥ 0}, tzn. $\phi_t(x)$ jest rozwiązaniem układu równań z warunkiem początkowym x(0) = x. Zauważmy, że dla równań Hamiltona pole A ma zerową dywergencję. Istotnie, mamy

div
$$A = \sum_{j=1}^{m} \frac{\partial A_j}{\partial x_j} = \sum_{i=1}^{N} \frac{\partial}{\partial p_i} \left(-\frac{\partial H}{\partial q_i} \right) + \frac{\partial}{\partial q_i} \left(\frac{\partial H}{\partial p_i} \right) = 0.$$

Okazuje się, że potok generowany przez pole wektorowe o zerowej dywergencji zachowuje objętość, tzn.

$$|A| = |\phi_t(A)|, \qquad t \ge 0.$$

Jest to tzw. Twierdzenie Liouville'a, znane z kursu równań różniczkowych zwyczajnych. Zatem potok równań Hamiltona zachowuje miarę Lebesguea. Ze względu na fakt, że energia E = H jest zachowana podczas ewolucji układu, układ w przestrzeni fazowej porusza się po zwartych poziomicach $\{H = E_0\}$. Na tych poziomicach również istnieje miara niezmiennicza, proporcjonalna do $\frac{dS}{|\nabla H|}$, gdzie dS jest miarą powierzchniową.

Podstawowym założeniem mechaniki statystycznej jest przyjecie, że unormowana do jedności miara $\mu_E = \frac{dS}{|\nabla H|}$ na powierzchni $\{H = E_0\}$ jest również miarą ergodyczną względem potoku Hamiltona, tzn. dla $A \subset \{H = E_0\}$ mamy

$$\mu(\phi_t(A)\Delta A) = 0 \implies \mu(A) = 0 \text{ lub } \mu(A) = 1.$$

Stwierdzenie to zostało udowodnione jedynie w kilku najprostrzych przypadkach.

Okazuje się, że dla niezmienniczej miary ergodycznej μ prawdziwa jest równość

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T f(\phi_t(x_0)) \, \mathrm{d}t = \int f \, \mathrm{d}\mu,$$

dla każdego punktu x_0 i dowolnej funkcji ciągłej f. Oczywiście pomijamy tutaj ścisłe sformułowanie tego twierdzenia. Jest to tzw. twierdzenie ergodyczne Birkhoffa. Równość ta jest niezwykle ważna, gdyż lewą stronę można zinterpretować jako wynik pomiaru wielkości f w naszym układzie. Faktycznie, mierząc pewną wielkość za pomocą urządzenia posiadającego bezwładność, nie mierzymy jej w określonej chwili czasowej. Otrzymujemy raczej uśrednienie czasowe tej wielkości.

Miara $\mu_E = \frac{dS}{|\nabla H|}$ na poziomicy stałej energii $\Gamma_E = \{H = E_0\} \cap \Gamma$ nazywana jest zespołem mikrokanonicznym. Jeśli $A : \Gamma_E \to \mathbb{R}$ jest pewną obserwablą (wielkością fizyczną), to mamy

$$\langle A \rangle_{mik} = \int_{\Gamma_E} A \, \mathrm{d}\mu_E.$$

Przykład 5. Prostym przykładem przekształcenia ergodycznego jest obrót o kąt niewymierny na okręgu. Niezmienniczą miarą ergodyczną jest tu miara Lebesguea. Mówiąc sciśle, rozważmy przekształcenie $T : [0,1) \rightarrow [0,1)$ zadane wzorem $T(x) = x + \alpha \mod 1$, gdzie α jest liczbą niewymierną. Wówczas dla dowolnej funkcji ciągłej $f : [0,1) \rightarrow \mathbb{R}$ prawdziwa jest równość

$$\lim_{n \to \infty} \frac{f(x) + f(T(x)) + \ldots + f(T^{n-1}(x))}{n} = \int_0^1 f(x) \, \mathrm{d}x.$$

Dowód tego faktu pozostawiamy jako zadanie. Wskazówka: wystarczy udowodnić powyższą równość dla funkcji $f_k(x) = e^{2\pi k i x}, k \in \mathbb{Z}.$

2.3 Zespół kanoniczny

Do tej pory zajmowaliśmy się zamkniętymi układami fizycznymi. Na układ nie oddziaływały żadne siły zewnętrzne. Bardzo często mamy jednak do czynienia z układem znajdującym się w pewnym otoczeniu o temperaturze T. Ale co to jest ta temperatura...? Zrozumienie tego pojęcia wymaga zapoznania się z koncepcjami termodynamiki fenomenologicznej. Na nasz użytek przyjmijmi, że układ znajdujący się w otoczeniu o stałej temperaturze T (takie otoczenie nazywamy czasem termostatem) nie ma ściśle określonej energii (układ w przestrzeni fazowej nie porusza się po poziomicach hamiltonianu), ale jego energia zmienia się w czasie ze względu na oddziaływanie z termostatem. Dla T = 0 układ po oddaniu całej możliwej energii do otoczenia znajduje się w położeniu o najniższej energii, czyli w położeniu w którym hamiltonian osiąga minimum. Jeśli T > 0, oddziaływanie z termostatem powoduje, że układ jest wytrącany z położenia o najmniejszej energii (zachodzą tzw. fluktuacje termicznych. Okazuje się, że prawdopodobieństwo znalezienia układu w stanie o energii H jest proporcjonalne do

$$e^{-\beta H}, \qquad \beta = \frac{1}{k_B T},$$

gdzie k_B jest stałą Boltzmanna. Odpowiednikiem miary μ_E jest teraz miara

$$\nu_T = \frac{e^{-\beta H}}{\int_{\Gamma} e^{-\beta H} \,\mathrm{d}\Gamma} \mathrm{d}\Gamma,$$

gdzie Γ jest miarą Lebesguea na przestrzeni fazowej Γ . Miara ta nazywana jest miarą Gibbsa lub zespołem kanonicznym. Mamy

$$\langle A \rangle_{kan} = \frac{\int_{\Gamma} A e^{-\beta H} \,\mathrm{d}\Gamma}{\int_{\Gamma} e^{-\beta H} \,\mathrm{d}\Gamma},$$

Widać wyraźnie, że jeśli $T \to 0^+$, preferowany jest stan układu odpowiadający minimu energii. Zobaczymy to wyraźniej w następnym podrozdziale. Jeśli $T \to \infty$, czynnik $e^{-\beta H}$ zbiega do 1 i otrzymujemy miarę jednostajną na Γ . Widzimy również, że dla skończonych T > 0 preferowane są położenia o niskich energiach.

2.4 Jednowymiarowy model Isinga

W tym podrozdziale rozważymy układ dyskretny o skończonej liczbie dostępnych stanów. W tym przypadku nie mamy dynamiki zadanej przez hamiltonian⁶. Mamy

 $^{^6\}mathrm{W}$ teorii procesów Markova w
prowadza się sztuczną dynamikę zwaną dynamiką Glaubera

jednak zamiar badać zachowanie układu, tzn. obliczać wartości średnie pewnych obserwabli, korzystając z zespołu kanonicznego.

Rozważmy układ N spinów s_1, s_2, \ldots, s_N , z których każdy może przyjmować wartość 1 lub -1. Jeśli $s_i = 1$, to *i*-ty spin jest ustawiony "do góry", a jeśli $s_i = -1$, to jest on ustawiony "w dół". Rozważmy następujący hamiltonian

$$H(s_1, \dots, s_n) = -J \sum_{i=1}^N s_i s_{i+1} - h \sum_{i=1}^N s_i,$$

gdzie przyjmujemy $s_{N+1} = s_1$. Pierwsza część opisuje energię oddziaływania między spinami. Stała J > 0 opisuje siłę tego oddziaływania i jest nazywana stałą sprzężenia. Stała h jest wartością zewnętrznego pola magnetycznego, w którym znajduje się układ. Możemy wyobrażać sobie, że spiny tworzą dyskretny okrąg i zachodzi oddziaływanie jedynie między sąsiednimi spinami. Minimum energii występuje wtedy, gdy wszystkie spiny są równoległe i ustawione zgodnie z polem magnetycznym.

Nasza przestrzeń fazowa składa się teraz ze wszystkich możliwych wyborów ustawień spinów. Jest to zatem zbiór $\{-1,1\}^N$. Zespół kanoniczny przyjmuje teraz postać dyskretnej miary na $\{-1,1\}^N$

$$\nu_T(\{s_1,\ldots,s_N\}) = \frac{e^{-\beta H(s_1,\ldots,s_N)}}{Z_N},$$

gdzie Z_N jest czynnikiem normalizującym zwanym suma statystyczną,

$$Z_N = \sum_{(s_1, \dots, s_N) \in \{-1, 1\}^N} e^{-\beta H(s_1, \dots, s_N)}$$

Rozważmy układ w temperaturze T = 0. Aby wyznaczyć zespoł kanoniczny w granicy $T \to 0^+$ zauważmy, że

$$\lim_{T \to 0^+} \nu_T(\{s_1, \dots, s_N\}) = \begin{cases} 1 & H(s_1, \dots, s_N) = \min_{s_1, \dots, s_N} H\\ 0 & H(s_1, \dots, s_N) \neq \min_{s_1, \dots, s_N} H \end{cases}$$

Zatem w T = 0 rozsądnie jest przyjąć, że zespół kanoniczny jest deltą Diraca w punkcie przestrzeni fazowej, w którym hamiltonian przyjmuje minimum.

Interesuje nas średnia magnetyzacja układu przypadająca na jeden spin, czyli

$$\langle m(N,T,h) \rangle_{kan} = \left\langle \frac{s_1 + \ldots + s_N}{N} \right\rangle_{kan}.$$

Jeśli h > 0 mamy m(N, T = 0, h) = 1, gdyż minimum hamiltonianu odpowiada konfiguracji $(1, \ldots, 1)$. Jeśli h < 0 to m(N, T = 0, h) = -1, bo tym razem minimum hamiltonianu odpowiada konfiguracji $(-1, \ldots, -1)$. Dla h = 0 mamy dwa minima hamiltonianu. Wracając do przypadku T > 0 widzimy, że dla h = 0 hamiltonian ma symetrię ze względu na odbicie spinów $(s_1, \ldots, s_N) \rightarrow (-s_1, \ldots, -s_N)$, czyli $H(s_1, \ldots, s_N) = H(-s_1, \ldots, -s_N)$. Zatem bez trudu otrzymujemy równość $\langle m(N, T, h = 0) \rangle_{kan} = 0$ dla wszystkich $T \ge 0$. Otrzymaliśmy, że m(N, T = 0, h) =sgnh. Funkcja m jest zatem nieciągłą funkcją zmiennej h. W szczególności nie jest to funkcja analityczna. W takiej sytuacji mówimy, że w układzie występuje przejście fazowe. Faktycznie, dla h < 0 mamy fazę (-) ze wszystkimi spinami ustawionymi do dołu, zaś dla h > 0 mamy fazę (+). Dla h = 0 mamy dwie konfiguracje minimalizujące energię. Taką sytuację nazywamy spontanicznym łamaniem symetrii – z fizycznego punktu widzenia układ wybiera jedną z dwóch faz i żadna z tych faz nie ma symetrii, którą ma hamiltonian.

Przypadek T = 0 jest nieco niefizyczny. Wykorzystaliśmy go jedynie dla prostego zilustrowania, czym jest przejście fazowe i spontaniczne łamanie symetrii. Wprowadźmy oznaczenia $K = \beta J$ oraz $H = \beta h$. Wówczas

$$Z_N(T,H) = \sum_{s_1,\dots,s_n \in \{-1,1\}} e^{K \sum_{i=1}^N s_i s_{i+1} + H \sum_{i=1}^N s_i}$$

Oczywiście dla T > 0 funkcja m(N, T, h) jest ciągłą, a nawet analityczna. Można się o tym przekonać zauważając, że

$$m(N,T,h) = \frac{1}{N} \frac{\mathrm{d}}{\mathrm{d}H} \ln Z_N(T,H) \tag{1}$$

oraz $Z_N(N, T, H)$ jest skończoną sumą funkcji analitycznych zmiennych T > 0 i H. Zatem w skończonych układach nie możemy obserwować przejść fazowych. Jednakże przejście fazowe może zajść w układzie po przejściu do granicy termodynamicznej. Oznacza to wzięcie granicy $N \to \infty$. Definiujemy gęstość energii swobodnej dla układu jednorodnego,

$$f(T,H) = -k_B T \lim_{N \to \infty} \frac{\ln Z_N(T,H)}{N}.$$

Analityczność tej funkcji oznacza brak przejścia fazowego w układzie nieskończonym.

Obliczymy magnetyzację w granicy termodynamicznej,

$$m(T, H) = \lim_{N \to \infty} m(N, T, H)$$

Pokażemy, że dla T > 0 funkcja m(T, H) jest ciągła. W szczególności mamy $\lim_{H\to 0} m(T, H) = 0$. Nasza strategia będzie bardzo prosta. Wyznaczymy sumę statystyczną $Z_N(T, H)$, a następnie obliczymy m(T, H, N) za pomocą wzoru (1). Na końcu bez trudu wyznaczymy f(T, H) i przekonamy się, że jest to funkcja analityczna zmiennych T, H.

Zastosujemy metodę macierzy przejścia. Niech

$$V(s_a, s_b) = \exp(Ks_a s_b + \frac{1}{2}h(s_a + s_b)), \quad s_a, s_b \in \{-1, 1\}.$$

Wyrażenie V możemy napisać w postaci macierzy

$$V = \left[\begin{array}{cc} e^{K-h} & e^{-K} \\ e^{-K} & e^{K+h} \end{array} \right]$$

Wówczas

$$Z_N = \sum_{s_1, \dots, s_N} V(s_1, s_2) V(s_2, s_3) \cdot \dots \cdot V(s_{N-1}, s_N) V(s_N, s_1).$$

Teraz pora na kluczową obserwację. Zauważmy, że

$$\sum_{s_2,...,s_N} V(i, s_2) V(s_2, s_3) \cdot \ldots \cdot V(s_{N-1}, s_N) V(s_N, j)$$

jest wyrazem (i, j) macierzy V^N . Zatem

$$\sum_{s_1,\dots,s_N} V(s_1,s_2) V(s_2,s_3) \cdot \dots \cdot V(s_{N-1},s_N) V(s_N,s_1) = \text{Tr} V^N.$$

Otrzymaliśmy wzór $Z_N(T, H) = \text{Tr}(V^N)$. Jeśli λ_-, λ_+ są wartościami własnymi macierzy V, to λ_-^N, λ_+^N są wartościami własnymi macierzy V^N , a zatem

$$Z_N(T,H) = \lambda_-^N + \lambda_+^N.$$

Wartości własne $\lambda_{-}^{N}, \lambda_{+}^{N}$ wyznaczamy bardzo łatwo znajdując miejsca zerowe wielomianu charakterystycznego macierzy V. Uważny czytelnik sprawdzi, że

$$\lambda_{-} = e^{K} \cosh H - (e^{2K} \sinh^{2} H + e^{-2K})^{1/2}, \quad \lambda_{+} = e^{K} \cosh H + (e^{2K} \sinh^{2} H + e^{-2K})^{1/2}.$$

Łatwo się przekonać, że $0 < \lambda_- < \lambda_+ < \infty$ oraz, że wartości własne są analitycznymi funkcjami T > 0 i H.

Mamy

$$\lim_{N \to \infty} m(N, T, H) = \lim_{N \to \infty} \frac{1}{N} \frac{\mathrm{d}}{\mathrm{d}H} \ln Z_N(T, H) = \lim_{N \to \infty} \frac{1}{N} \frac{\frac{\partial Z_N(T, H)}{\partial H}}{Z_N(T, H)}$$
$$= \lim_{N \to \infty} \frac{\lambda_-^{N-1} \frac{\partial \lambda_-}{\partial H} + \lambda_+^{N-1} \frac{\partial \lambda_+}{\partial H}}{\lambda_-^N + \lambda_+^N} = \frac{1}{\lambda_+} \frac{\partial \lambda_+}{\partial H}.$$

Po krótkich obliczeniach otrzymujemy

$$m(T,H) = \frac{e^{K} \sinh H}{\left(e^{2K} \sinh^{2} H + e^{-2K}\right)^{1/2}}$$

Oczywiście funkcja m(T, H) jest ciągła, m(T, 0) = 0 oraz $\lim_{H\to\infty} m(T, H) = 1$, $\lim_{H\to-\infty} m(T, H) = -1$. Oznacza to, że silne pole wymusza ustawianie się spinów zgodnie z jego kierunkiem.

Zauważmy, że

$$\ln \lambda_{+} = \frac{1}{N} \ln(\lambda_{+}^{N}) \leq \frac{1}{N} \ln(\lambda_{-}^{N} + \lambda_{+}^{N}) \leq \frac{1}{N} \ln(2\lambda_{+}^{N}) = \frac{1}{N} \ln 2 + \ln \lambda_{+} \xrightarrow[N \to \infty]{} \ln \lambda_{+}.$$

Stąd

$$f(H,T) = -k_B T \ln \left(e^K \cosh H + (e^{2K} \sinh^2 H + e^{-2K})^{1/2} \right).$$

Jest to funkcja jawnie analityczna.

Rozważymy teraz przypadek dodatnich warunków brzegowych. Rozważ
my układ ${\cal N}$ spinów z hamiltonianem

$$H(s_1, \dots, s_n) = -J \sum_{i=1}^{N-1} s_i s_{i+1} - h \sum_{i=1}^{N} s_i,$$

przy czym $s_1 = s_N = 1$. Tym razem mamy

$$Z_N = \sum_{(s_2,\dots,s_{N-1})\in\{-1,1\}^N} e^{-\beta H(s_1=1,s_2,\dots,s_{N-1},s_N=1)}.$$

Zauważmy, że

$$Z_N = \sum_{s_2,\dots,s_{N-1}} V(1,s_2) V(s_2,s_3) \cdot \dots \cdot V(s_{N-2},s_{N-1}) V(s_{N-1},1) = (V^N)_{11}.$$

Aby obliczyć sumę statystyczną musimy zatem przedstawić macier
z ${\cal V}$ w postaci

$$V = U \left[\begin{array}{cc} \lambda_{-} & 0\\ 0 & \lambda_{+} \end{array} \right] U^{-1}.$$

Wówczas

$$V^N = U \begin{bmatrix} \lambda_-^N & 0\\ 0 & \lambda_+^N \end{bmatrix} U^{-1}.$$

Stąd

$$Z_N = (V^N)_{11} = w_1(K, H)\lambda_-^N + w_2(K, H)\lambda_+^N,$$

gdzie w_1, w_2 są pewnymi współczynnikami, które można wyznaczyć, znajdując macierz U. Polecamy to ćwiczenie miłośnikom GALu. Konkluzja jest następująca. Mamy

$$m(N,T,H) = \frac{1}{N} \frac{\mathrm{d}}{\mathrm{d}H} \ln \left(w_1(K,H)\lambda_-^N + w_2(K,H)\lambda_+^N \right).$$

Podobnie jak poprzednio bez trudu otrzymujemy

$$\lim_{N \to \infty} m(N, T, H) = \frac{1}{\lambda_+} \frac{\partial \lambda_+}{\partial H},$$

a zatem przyjęcie dodatnich warunków nie wymusza zmiany średniej magnetyzacji przypadającej na jeden spin w granicy termodynamicznej. W szczególności w granicy termodynamicznej otrzymanej poprzez zadanie dodatnich warunków brzegowych w zerowym polu zewnętrznym średnia magnetyzacja jest równa 0.

Okazuje się, że w przypadku dwuwymiarowego modelu Isinga sytuacja wygladą inaczej. Tutaj zadanie dodatnich warunków brzegowych będzie miało wpływ na średnią magnetyzację w granicy termodynamicznej.

Przykład 6. Rozważmy jednowymiarowy model Isinga ze swobodnymi warunkami brzegowymi z z hamiltonianem

$$H(s_1, \dots, s_n) = -J \sum_{i=1}^{N-1} s_i s_{i+1} - h \sum_{i=1}^{N} s_i.$$

Wyznacz sumę statystyczną dla tego układu i oblicz średnią magnetyzację.

2.5 Dwuwymiarowy model Isinga. Argument Peierlsa.

Rozważmy kwadratową podsieć $D = \{0, 1, \dots, N, N+1\}^2$ kraty całkowitoliczbowej \mathbb{Z}^2 . Zadajmy dodatnie warunki brzegowe, tzn. spiny w węzłach sieci ze zbioru

$$\partial D = \{(x, y) \mid x \in \{0, N+1\} \text{ lub } y \in \{0, N+1\}\}$$

mają wartość +1. Hamiltonian układu jest równy

$$H_{+} = -J \sum_{\substack{i, j \in D, \\ |i-j| = 1}} s_i s_j - h \sum_{i \in D} s_i.$$

Zauważmy, że zadanie dodatnich warunków brzegowych nie ma wpływu na gęstość energii swobodnej w granicy termodynamicznej. Istotnie, jeśli chcemy porównać hamiltonian H_+ z hamiltonianem

$$H = -J \sum_{\substack{i, j \in D \setminus \partial D, \\ |i-j| = 1}} s_i s_j - h \sum_{i \in D \setminus \partial D} s_i,$$

odpowiadającym swobodnym warunkom brzegowym, otrzymamy

$$H - 4(J+h)\sqrt{|D|} \le H_+ \le H + 4(J+h)\sqrt{|D|}$$

dla dowolnej konfiguracji spinów na kracie $D \setminus \partial D$. Wynika stąd nierówność dla sum statystycznych

$$Z_N e^{-4\beta(J+h)\sqrt{|D|}} \le Z_N^+ \le Z_N e^{+4\beta(J+h)\sqrt{|D|}}.$$

Definiujemy gęstości energii swobodnych,

$$f_N = -\frac{1}{\beta |D \setminus \partial D|} \ln Z_N, \qquad f_N^+ = -\frac{1}{\beta |D|} \ln Z_N^+.$$

otrzymujemy

$$f_N \sqrt{\frac{|D \setminus \partial D|}{|D|} - \frac{4(J+h)}{\sqrt{|D|}}} \le f_N^+ \le f_N \sqrt{\frac{|D \setminus \partial D|}{|D|} + \frac{4(J+h)}{\sqrt{|D|}}}.$$

Zauważmy, że $|D| = (N+2)^2$. Biorąc $N \to \infty$ otrzymujemy lim $\sup_{N\to\infty} f_N^+ = \lim \sup_{N\to\infty} f_N$. W granicy termodynamicznej rozważamy granicę górną, ponieważ chcemy tu pominąć kwestię istnienia pełnej granicy.

Rozważmy układ z dodatnimi warunkami brzegowymi. Wprowadzamy pojęcie kraty dualnej, tzn.

$$\tilde{D} = \left\{\frac{1}{2}, \frac{3}{2}, \dots, N + \frac{1}{2}\right\}^2.$$



Rysunek 8: Przykładowa konfiguracja spinów i krawędzie na kracie dualnej.

Dwa sąsiednie wierzchołki na kracie \tilde{D} są połączone krawędzią, jeśli krawędź normalnej kraty tę krawędź przecinająca łączy spiny o przeciwnym znaku. Na Rysunku 8 pokazano przykładową konfigurację spinów wraz z krawędziami na kracie dualnej.

Zauważmy, że każdy wierzchołek na kracie dualnej ma stopień 0, 2 lub 4 (tu kluczowe są dodatnie warunki brzegowe!). Graf, którego każdy wierzchołek ma stopień parzysty można podzielić na krawędziowo rozłączne cykle. W tym celu bierzemy dowolny wiecholek o niezerowym stopniu i idziemy dowolnie po jeszcze nie użytych krawędziach aż do momentu ponownego spotkania z wierzchołkiem startowym. Procedura ta jest wykonalna ze względu na parzystość stopni wierzchołków – jeśli do jakiegoś wierzchołka trafimy, to zawsze możemy z niego wyjść, chyba, że trafiliśmy do początkowego wierzchołka. Otrzymany cykl usuwamy. Otrzymany graf ma nadal parzyste stopnie wierzcołków. Następnie powtarzamy procedurę aż do wyczerpania wszystkich krawędzi.

Jeśli dodatkowo zażądamy, aby zawsze wykonywać skręt w lewo, jeśli jest na skrzyżowaniu taka możliwość, otrzymamy konfigurację pętli, które się wzajemnie nie przecinają i nie mają samoprzecięć (przecięcie to sytuacja, w które w wierzchołku stopnia 4 spotykają się dwie drogi idące przez skrzyżowanie na wprost – jeśli skręcamy zawsze w lewo, to przez tego typu skrzyżowanie nigdy na wprost nie przejdziemy). Oczywiście wybór pętli które się wzajemnie nie przecinają i nie mają samoprzecięć nie jest jednoznaczny (patrz Rysunek 9). Wynika to z faktu, że przez wierzchołek stopnia 4 możemy przejść na dwa sposoby – skręcić w lewo lub w prawo. Jeśli jednak zawsze skręcamy w lewo, ta niejednoznaczność zostaje usunięta. Łatwo też zauważyć, że zbiór konturów zadaje w sposób jednoznaczny konfigurację spinów.

Każdej konfiguracji spinów odpowiada pewien zbiór rozłącznych konturów. Kontury mogą mieć długości 4,6,8,.... Przypuśmy, że jest m(l) konturów długości l. Zauważmy, że

$$m(l) \le |D| \cdot 4 \cdot 3^{l-1} \cdot \frac{1}{2l} \le \frac{3^{l}|D|}{l}$$

Wynika to z faktu, że kontur możemy budować odcinkami. Na mniej niż |D| sposobów wybieramy wierzchołek początkowy. Następnie możemy iść w czterech możliwych kierunkach. W kolejnym kroku mamy już do wyboru tylko 3 kierunki, bo nie możemy się wracać. Oczywiście nie dbamy w tej chwili o to, czy kontur się zamknie, czy nie. Jeśli jednak nasza ścieżka się zamknęła, to można było ją uzyskać na 2l sposobów (biorąc różne punkty początkowe i uwzględniając możliwe dwa kierunki obiegania konturu). Stąd czynnik 1/2l.

Niech ω oznacza pewną konfigurację spinów. Odpowiada jej pewna konfiguracja konturów. Niech $X_l^{(i)}$, $i = 1, 2, \ldots, m(l)$ będzie zdefiniowane następująco,

$$X_l^{(i)}(\omega) = \begin{cases} 1 & \text{jeśli } i\text{-ty kontur o długości } l \text{ występuje w konfiguracji } \omega \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

Niech $N_{-}(\omega)$ oznacza liczbę ujemnych spinów w konfiguracji ω . Każdy ujemny spin jest otoczony przez pewien kontur. Liczba spinów zawarta wewnątrz konturu długości l nie przekracza $\frac{l^2}{16}$. Zatem

$$N_{-}(\omega) \le \sum_{l=4,6,8,\dots} \frac{l^2}{16} \sum_{i=1}^{m(l)} X_l^{(i)}(\omega).$$

Zauważmy, że

$$\left\langle X_{l}^{(i)} \right\rangle_{kan} = \frac{\sum_{\omega:X_{l}^{(i)}(\omega)=1} e^{-\beta H(\omega)}}{\sum_{\omega} e^{-\beta H(\omega)}}.$$

Jeśli kontur γ odpowiadający zmiennej $X_l^{(i)}$ występuje w konfiguracji ω , to odwracając wszystkie spiny wewnątrz tego konturu otrzymujemy konfigurację ω^* , w której już kontur γ nie występuje. Inne kontury pozostają bez zmian. Energia tak otrzymanej konfiguracji różni się od konfiguracji początkowej o -2Jl, gdzie l oznacza długość konturu γ . Czyli,

$$H(\omega^{\star}) = H(\omega) - 2Jl.$$

Mamy

$$\left\langle X_l^{(i)} \right\rangle_{kan} = \frac{\sum_{\omega: X_l^{(i)}(\omega)=1} e^{-\beta H(\omega)}}{\sum_{\omega} e^{-\beta H(\omega)}} \le \frac{\sum_{\omega: X_l^{(i)}(\omega)=1} e^{-\beta H(\omega)}}{\sum_{\omega^*: X_l^{(i)}(\omega)=1} e^{-\beta H(\omega)}} = e^{-2J\beta l}.$$

Stąd

$$\langle N_{-} \rangle_{kan} \leq \sum_{l=4,6,8,\dots} \frac{l^2}{16} \sum_{i=1}^{m(l)} \left\langle X_l^{(i)} \right\rangle_{kan} \leq \sum_{l=4,6,8,\dots} \frac{l^2}{16} \sum_{i=1}^{m(l)} e^{-2Jl\beta} = \sum_{l=4,6,8,\dots} \frac{l^2}{16} e^{2J\beta l} m(l)$$

$$\leq \sum_{l=4,6,8,\dots} \frac{l^2}{16} e^{-2J\beta l} \frac{3^l |D|}{l} = \frac{|D|}{16} \sum_{l=4,6,8,\dots} l(3e^{-2J\beta})^l.$$

Niech $x = (3e^{-2J\beta})^2$. Mamy

$$\sum_{l=4,6,8,\dots} l(3e^{-2J\beta})^l = \sum_{l=4,6,8,\dots} lx^{l/2} = 2x(2x+3x^2+4x^3\dots) = 2x(x^2+x^3+\dots)'$$
$$= 2x\left(\left(\frac{1}{1-x}\right)'-1\right) = 2x\left(\frac{1}{(1-x)^2}-1\right).$$

Zauważmy, że jeśli β jest duże, czyli x jest małe, mamy

$$\langle N_{-} \rangle_{kan} \leq \frac{|D|}{4},$$

czyli

$$m_+(T, H = 0, N) = \frac{1}{|D|} \langle N_+ - N_- \rangle_{kan} = 1 - \frac{2}{|D|} \langle N_- \rangle_{kan} \ge \frac{1}{2}.$$

Oznacza to, że dla dostatecznie niskich temperatur, niezależnie od wartości N = |D| mamy dodatnią, oddzieloną od zera magnetyzację.

Zauważmy na koniec, że funkcja f w granicy termodynamicznej nie jest różniczkowalna. Funkcje f oraz f_N są funkcjami wklęsłymi zmiennej h. Istotnie, każda funkcja postaci

$$h \mapsto -\ln\left(\int g(x)e^{hf(x)} d\mu(x)\right)$$

jest wklęsła. Udowadniamy to poprzez dwukrotne różniczkowanie (zadanie!).



Rysunek 9: Wybór pętli nie jest jednoznaczny.

Mamy

$$\begin{aligned} \frac{f(h) - f(0)}{h} &\leq \sup_{h>0} \frac{f(h) - f(0)}{h} = \sup_{h>0} \liminf_{N\to\infty} \frac{f_N^+(h) - f_N^+(0)}{h} \\ &\leq \liminf_{N\to\infty} \sup_{h>0} \frac{f_N^+(h) - f_N^+(0)}{h} = \liminf_{N\to\infty} (f_N^+)'(0) \\ &= -\liminf_{N\to\infty} m_+(T, H = 0, N) \leq -\frac{1}{2}. \end{aligned}$$

Z symetrii mamy f(-h) = f(h), a zatem granica prawostronna ilorazów różnicowych funkcji f w pukcie h = 0 jest nie większa niż $-\frac{1}{2}$, zaś granica lewostronna jest nie mniejsza niż $\frac{1}{2}$. Zatem f nie jest różniczkowalna w punkcie h = 0, a co za tym idzie w układzie występuje przejście fazowe. Ponadto nierówność

$$\liminf_{N \to \infty} m_+(T, H = 0, N) \ge \frac{1}{2}$$

oznacza, że dla niezerowych temperatur mamy spontaniczną magnetyzację.

3 Perkolacje i grafy losowe

3.1 Model Erdösa-Rényi, monotoniczne funkcje boolowskie

Rozważmy następujący model grafu losowego. Przypuśćmy że mamy dane N wierzchołków. Mamy zatem $n = \binom{N}{2}$ par wierzchołków. Dla każdej pary wierzchołków (niezależnie!) rysujemy krawędź z prawdopodobieństwem p. Z prawdopodobieństwem 1-p wierzchołki te nie są połączone krawędzią. Zatem, dla przykładu, prawdopodobieńswo, że nasz graf losowy G(N, p) jest kliką K_N jest równe p^n . Opisany model pochodzi od Erdösa i Rényi, [ER].

Będą nas interesować zdarzenia monotoniczne dla grafu G(N, p). Zdarzenie jest monotoniczne, jeśli dołożenie krawędzi w grafie nie może spowodować, że zdarzenie przestanie zachodzić. Przykładem zdarzenia monotonicznego jest zdarzenie graf G(N,p) jest spójny. Oczywiście, jeśli dla pewnego wyboru krawędzi nasz graf jest spójny, to po dołożeniu nowej krawędzi graf nadal będzie spójny.

Opiszemy teraz przestrzeń probabilistyczną dla modelu Erdösa-Rényi. Przestrzenią tą będzie zbiór $\{-1,1\}^n$ z σ -ciałem wszystkich podzbiorów i miarą produktową $\mu_p^n = ((1-p)\delta_{-1} + p\delta_1)^{\otimes n}$, gdzie $n = \binom{N}{2}$ jest liczbą wszystkich par wierzchołków grafu. Współrzędne wektora ze zbioru $\{-1,1\}^n$ odpowiają parom wierzchołków, przy czym liczba 1 oznacza, że między wierzchołkami jest krawędź, a liczba -1 oznacza brak krawędzi. Zatem klice K_N odpowiada teraz wierzchołek $(1, 1, \ldots, 1)$.

Zdarzenia losowe to podzbiory naszej przestrzeni probabilistycznej. Zdarzenie graf G(N,p) jest kliką K_N jest zbiorem jednoelementowym. Zdarzenie graf G(N,p)zawiera przynajmniej k krawedzi jest zbiorem złożonym ze wszystkich punktów, które mają przynajmniej k współrzędnych równych 1. Prawdopodobieństwo zdarzenia $A \subseteq \{-1,1\}^n$ jest równe

$$\mathbb{P}_p(A) = \sum_{(x_1, \dots, x_n) \in A} p^{|\{i: x_i=1\}|} (1-p)^{|\{i: x_i=-1\}|}.$$
(2)

Oczywiście $\mathbb{P}_p(A) = \mu_p^n(A).$

Zostawmy na moment w spokoju model Erdösa-Rényi i zajmijmy się dowolnymi zdarzeniami na przestrzeni probabilistycznej $\{-1, 1\}^n$ z miarą μ_p^n , gdzie *n* jest pewną liczbą naturalną. Dowolne zdarzenie $A \subset \{-1, 1\}^n$ możemy identyfikować z pewną funkcją $f : \{-1, 1\}^n \to \{-1, 1\}$. Takie funkcje nazywamy funkcjami boolowskimi. Konkretnie, dla podzbioru A definiujemy funkcję $f = f_A$ wzorem $f_A(x) = \mathbf{1}_A(x) - \mathbf{1}_{A^c}(x)$. Oczywiście $\mathbb{P}_p(A) = \mathbb{P}_p(f_A = 1)$. Rozważmy następujący przykład. Niech *n* będzie liczbą nieparzystą. Przypuśćmy, że społeczeństwo złożone z *n* obywateli musi wybrać prezydenta. Mamy dwóch kandydatów, Freda i Barniego. Przyjmujemy, że $x_i = 1$ oznacza, że *i*-ty głosujący głosuje na Freda, zaś $x_i = -1$ oznacza, że jego faworytem jest Barney. Metody wyboru kandydata mogą być różne. Każda taka metoda jest pewną funkcją boolowską. Jeśli $x = (x_1, \ldots, x_n)$ jest wektorem oddanych głosów i f(x) = 1, to wygrywa Fred. Jeśli f(x) = -1, zwycięzcą jest Barney. Dla przykładu, jeśli mamy do czynienia z dyktaturą, to o wszystkim decyduje tylko jeden głosujący (dyktator). Wtedy funkcja boolowska, na mocy które obliczany jest wynik wyborów zadana jest wzorem

$$\operatorname{Dict}_n(x_1,\ldots,x_n)=x_1.$$

Przypuśćmy jednak, że nasz ustrój jest oparty na demokracji. Jest to związane z pewną symetrią – głos każdego obywatela jest tak samo ważny. Decyduje większąć, a zatem tym razem nasza funkcja dana jest wzorem

$$\operatorname{Maj}_n(x_1,\ldots,x_n) = \operatorname{sgn}(x_1+\ldots+x_n).$$

Nasza funkcja jest symetryczna ze względu na permutację wspołrzędnych, mianowicie

$$\operatorname{Maj}_n(x_{\sigma(1)},\ldots,x_{\sigma(n)}) = \operatorname{Maj}_n(x_1,\ldots,x_n)$$

dla dowolnej permutacji σ zbioru $\{1, \ldots, n\}$. Ogólniej, zdarzenie $\{f = 1\}$ nazwiemy symetrycznym, jeśli istnieje tranzytywna grupa permutacji Γ działająca na $\{1, \ldots, n\}$ taka, że zbiór $\{f = 1\}$ jest niezmienniczy względem działania tej grupy. Dla symetrycznego zdarzenia grafowego grupą Γ jest grupa permutująca wierzchołki grafu. Ponadto, funkcja Maj_n jest monotoniczna (funkcja $f : \{-1, 1\}^n \to \{-1, 1\}$ jest monotoniczna jeśli $x_i \leq y_i$ dla wszystkich $1 \leq i \leq n$ implikuje $f(x_1, \ldots, x_n) \leq f(y_1, \ldots, y_n)$). Niech $\theta_f(p) = \mathbb{P}_p(f = 1)$. Udowodnimy następujący, intuicyjnie oczywisty lemat.

Lemat 1. Niech $f : \{-1,1\}^n \to \{-1,1\}$ będzie funkcją monotoniczną. Wówczas $\theta_f(p) = \mathbb{P}_p(f=1)$ jest funkcją niemalejącą.

Dowód. Zamiast przestrzeni probabilistycznej $\{-1,1\}^n$ z miarą μ_p^n rozważmy przestrzeń $[0,1]^n$ z σ -ciałem zbiorów mierzalnych i miarą Lebesguea λ_n . Definiujemy fukcję

$$\hat{f}_p(x_1,\ldots,x_n) = f(\mathbf{1}_{[0,p]} - \mathbf{1}_{(p,1]}(x_1),\ldots,\mathbf{1}_{[0,p]} - \mathbf{1}_{(p,1]}(x_1)).$$

Wówczas jeśli $p_1 \leq p_2$, to

$$\theta_f(p_1) = \mathbb{P}_{p_1}(f=1) = \lambda_n(\tilde{f}_{p_1}=1) \le \lambda_n(\tilde{f}_{p_2}=1) = \mathbb{P}_{p_1}(f=1) = \theta_f(p_2),$$

bo $\{\tilde{f}_{p_1}=1\} \subset \{\tilde{f}_{p_2}=1\}.$

Zastosowaną technikę nazywamy powiązaniem miar. Zamiast rozważać ustaloną funkcję na różnych przestrzeniach probabilistycznych, przeszliśmy do obrazu, w którym przestrzeń probabilistyczna jest ustalona, ale funkcje są różne.

Ze wzoru (2) wynika natychmiast następująca uwaga.

Uwaga 1. Niech $f : \{-1,1\}^n \to \{-1,1\}$. Wówczas $\theta_f(p) = \mathbb{P}_p(f=1)$ jest funkcją ciągłą.

Ustalmy $\varepsilon \in (0, \frac{1}{2})$. Dla funkcji monotonicznej $f\{-1, 1\}^n \to \{-1, 1\}$ definiujemy $p_-(\varepsilon)$ i $p_+(\varepsilon)$ poprzez relacje $\theta_f(p_-(\varepsilon)) = \varepsilon$ oraz $\theta_f(p_+(\varepsilon)) = 1 - \varepsilon$. Okazuje się, że przy dodatkowym założeniu symetrii funkcji f, przedział $(p_-(\varepsilon), p_+(\varepsilon))$ ma dla dużych n bardzo małą długość. Zanim przejdziemy do ścisłego sformułowania, zilustrujemy to twierdzenie przykładem.

Przykład 7. Rozważmy funkcję $\operatorname{Maj}_n(x_1, \ldots, x_n) = \operatorname{sgn}(x_1 + \ldots + x_n)$, gdzie *n* jest liczbą nieparzystą. Naszym celem jest oszacowanie $p_{-}(\varepsilon)$ i $p_{+}(\varepsilon)$. Będzie nam potrzebny lemat dotyczący schematu Bernoulliego.

Lemat 2 (Nierówność Bernsteina). Niech S_n będzie liczbą sukcesów w schemacie Bernoulliego z prawdopodobieństwem sukcesu p. Wówczas

$$\mathbb{P}_p\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \le 2e^{-2n\varepsilon^2}.$$
(3)

Zauważmy, że $p_{-}(\varepsilon) < \frac{1}{2}$. Korzystając z nierówności (3) otrzymujemy

$$\varepsilon = \theta_f(p_-(\varepsilon)) = \mathbb{P}_{p_-(\varepsilon)} \left(x_1 + \ldots + x_n > 0 \right) = \mathbb{P}_{p_-(\varepsilon)} \left(\frac{S_n}{n} > \frac{1}{2} \right)$$
$$= \mathbb{P}_{p_-(\varepsilon)} \left(\frac{S_n}{n} > p_-(\varepsilon) + \frac{1}{2} - p_-(\varepsilon) \right) \le 2e^{-2n(\frac{1}{2} - p_-(\varepsilon))^2}.$$

Zatem

$$\frac{1}{2} \ge p_{-}(\varepsilon) \ge \frac{1}{2} - \sqrt{\frac{\ln(2/e)}{2n}}.$$

Oczywiście $p_+(\varepsilon) = 1 - p_-(\varepsilon)$. Otrzymujemy stąd

$$|p_+(\varepsilon) - p_-(\varepsilon)| \le 2\sqrt{\frac{\ln(2/e)}{2n}} \xrightarrow[n \to \infty]{} 0.$$
Dowód nierówności Bernsteina, [JS]. Zauważmy najpierw, że dla $p,q \ge 0$ takich, że p+q=1 prawdziwa jest nierówność

$$pe^{qx} + qe^{-px} \le e^{\frac{1}{8}x^2}, \quad x \in \mathbb{R}.$$
(4)

Aby udowodnić tę nierówność zauważmy najpierw, że zamieniając ewentualnie p i q rolami możemy bez zmniejszenia ogólności zakładać, że $x \ge 0$. Niech $g(x) = \ln(pe^{qx} + qe^{-px})$. Mamy wykazać nierówność $g(x) \le \frac{1}{8}x^2$. Zauważmy, że g(0) = g'(0) = 0. Zatem wystarczy udowodnić, że $g''(x) \le \frac{1}{4}$ dla $x \ge 0$. Wynika to z nierówności

$$g''(x) = \left(\frac{pqe^{qx} - pqe^{-px}}{pe^{qx} + qe^{-px}}\right)' = pq\left(\frac{e^x - 1}{pe^x + q}\right)' = pq\frac{e^x(pe^x + q) - pe^x(e^x - 1)}{(pe^x + q)^2}$$
$$= \frac{pqe^x}{(pe^x + q)^2} = \frac{pqe^x}{(pe^x - q)^2 + 4pqe^x} \le \frac{1}{4}.$$

Niech λ będzie dowolną liczbą dodatnią. Korzystając z udowodnionej powyżej nierówności otrzymujemy

$$\mathbb{P}_p\left(\frac{S_n}{n} > p + \varepsilon\right) = \sum_{k > n(p+\varepsilon)} \binom{n}{k} p^k (1-p)^{n-k}$$

$$\leq \sum_{k > n(p+\varepsilon)} \binom{n}{k} p^k (1-p)^{n-k} e^{-\lambda(n(p+\varepsilon)-k)}$$

$$\leq \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} e^{-\lambda(n(p+\varepsilon)-k)}$$

$$= e^{-\lambda n\varepsilon} \sum_{k=0}^n \binom{n}{k} (pe^{\lambda q})^k (qe^{-\lambda p})^{n-k}$$

$$= e^{-\lambda n\varepsilon} (pe^{\lambda}q + qe^{-\lambda p})^n \leq e^{-\lambda n\varepsilon} e^{\frac{1}{8}\lambda^2 n}.$$

Przyjmując $\lambda=4\varepsilon$ trzymujemy

$$\mathbb{P}_p\Big(\frac{S_n}{n} > p + \varepsilon\Big) \le e^{-2n\varepsilon^2}.$$

Nierówność

$$\mathbb{P}_p\left(\frac{S_n}{n}$$

można łatwo otrzymać z powyższej nierówności zamieniając rolami p i q. Wówczas S_n odpowiada wyrażenie $n - S_n$.

Okazuje się, że dla dowolnej monotonicznej funkcji boolowskiej $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$, dla której zdarzenie $\{f = 1\}$ jest symetryczne, wyrażenie $p_+(\varepsilon) - p_{(\varepsilon)}$ dąży do 0 dla $n \rightarrow \infty$. Konkretnie, prawdziwe jest następujące oszacowanie.

Twierdzenie 1 (Kalai, Friedgut). Niech $\varepsilon \in (0, \frac{1}{2})$ i niech $p_{-}(\varepsilon)$ i $p_{+}(\varepsilon)$ spełniają $\theta_{f}(p_{-}(\varepsilon)) = \varepsilon$ oraz $\theta_{f}(p_{+}(\varepsilon)) = 1 - \varepsilon$. Załóżmy ponadto, że f jest funkcją monotoniczną i $\{f = 1\}$ jest zdarzeniem symetryczną. Wówczas

$$p_{+}(\varepsilon) - p_{-}(\varepsilon) \le c_1 \frac{\ln(1/2\varepsilon)}{\ln n}.$$
 (5)

W szczególności, dla symetrycznych zdarzeń w modelu Erdösa-Rényi (czyli zdarzeń niezmienniczych ze względu na przenumerowanie krawędzi) mamy następujące twierdzenie.

Twierdzenie 2. Rozważmy graf losowych G(N, p) i niech A będzie monotonicznym zdarzeniem symetrycznym. Niech $\varepsilon \in (0, \frac{1}{2})$ i niech $p_{-}(\varepsilon)$ i $p_{+}(\varepsilon)$ spełniają $\mathbb{P}_{p_{-}(\varepsilon)} = \varepsilon$ oraz $\mathbb{P}_{p_{+}(\varepsilon)} = 1 - \varepsilon$. Wówczas

$$p_+(\varepsilon) - p_-(\varepsilon) \le c_1 \frac{\ln(1/2\varepsilon)}{\ln \binom{N}{2}}.$$

Zanim przejdziemy do dowodu powyższych twierdzeń, musimy udowodnić kilka faktów z zakresu tzw. spektralnej teorii funkcji boolowskich.

3.2 Spektralna teoria funkcji boolowskich

3.2.1 Układ Walsha-Fouriera

Przez pewien czas ograniczymy się do analizy przypadku $p = \frac{1}{2}$. Miara $\mu^n = \mu_{1/2}^n$ jest rozłożona jednostajnie na zbiorze $\{-1, 1\}^n$. Czasem, dla wygody, będzie pisać μ zamiast μ^n , jeśli będzie jasne, z jakim wymiarem n aktualnie pracujemy. Zatem dla $A \subseteq \{-1, 1\}^n$ mamy

$$\mu^{n}(A) = \mathbb{P}_{1/2}(A) = \frac{1}{2^{n}}|A|,$$

gdzie |A|oznacza liczbę elementów zbioruA. Wartość oczekiwana funkcji $f:\{-1,1\}^n\to\mathbb{R}$ dana jest wzorem

$$\mathbb{E}f = \frac{1}{2^n} \sum_{x \in \{-1,1\}^n} f(x).$$

Dla $p \ge 1$ możemy również zdefiniować p-tą normę funkcji f,

$$||f||_p = (\mathbb{E}|f|^p)^{1/p}$$

Funkcje określone na zbiorze $\{-1,1\}^n$ tworzą przestrzeń liniową wymiaru 2ⁿ. Funkcje charakterystyczne punktów,

$$\delta_y(x) = \begin{cases} 1 & x = y \\ 0 & x \neq y \end{cases},$$

tworzą bazę ortogonalną tej przestrzeni. Wprowadzamy strukturę przestrzeni Hilberta $L_2(\{-1,1\}^n,\mu^n)$ za pomocą iloczynu skalarnego

$$\langle f,g\rangle = \mathbb{E}fg = \frac{1}{2^n} \sum_{x \in \{-1,1\}^n} f(x)g(x).$$

W naszych zastosowaniach potrzebna będzie jednak inna baza, zwana układem Walsha-Fouriera. Przyjmujemy oznaczenie $[n] = \{1, \ldots, n\}$ i dla $S \subseteq [n]$ definiujemy funkcję

$$w_S(x_1,\ldots,x_n) = \prod_{i\in S} x_i, \quad S\subset [n], \quad w_{\emptyset} \equiv 1.$$

Łatwo zauważamy, że $w_S \cdot w_T = w_{S\Delta T}$. Miara μ^n jest miarą produktową, a zatem

$$\mathbb{E}x_{i_1}\cdot\ldots\cdot x_{i_k}=\mathbb{E}x_{i_1}\cdot\ldots\cdot\mathbb{E}x_{i_k}=0.$$

Wynika stąd ortonormalność układu Walsha-Fouriera,

$$\mathbb{E}w_S = \left\{ \begin{array}{cc} 1 & S = \emptyset \\ 0 & S \neq \emptyset \end{array} \right\}, \qquad \mathbb{E}w_{S\Delta T} = \left\{ \begin{array}{cc} 1 & S = T \\ 0 & S \neq T \end{array} \right\}.$$

Każda funkcja $f:\{-1,1\}^n \to \mathbb{R}$ może być zatem przedstawiona w postaci sumy

$$f = \sum_{S \subseteq [n]} a_S w_S,$$

gdzie $(a_S)_{S \subseteq [n]}$ są pewnymi liczbami rzeczywistymi. Liczby te nazywamy czasem spektrum funkcji f i piszemy $a_S = \hat{f}(S)$. Zauważmy, że

$$\langle f, w_T \rangle = \left\langle \sum_{S \subset [n]} a_s w_S, w_T \right\rangle = \sum_{S \subset [n]} a_S \langle w_S, w_T \rangle = a_T,$$

Możemy więc napisać

$$f = \sum_{S \subset [n]} \hat{f}(S) w_S = \sum_{S \subset [n]} \langle f, w_S \rangle w_S.$$

3.2.2 Hiperkontrakcja

W tym podrozdziale udowodnimy nierówność, która okaże się być kluczową w dowodzie Twierdzenia 1. Przedstawiona poniżej własność układu Walsha-Fouriera nosi nazwę hiperkontrakcji.

Twierdzenie 3 ([Bo], [Be], [G1]). Dla dowolnych liczb rzeczywistych $(a_S)_{S\subseteq[n]}$ i dowolnej liczby $\delta \in [0, 1]$ prawdziwa jest nierówność

$$\left\|\sum_{S\subseteq [n]} a_S w_S \delta^{|S|}\right\|_2 \le \left\|\sum_{S\subseteq [n]} a_S w_S\right\|_{1+\delta^2}$$

Zanim przystąpimy do dowodu tego twierdzenia, wprowadzimy ważne pojęcie operatorów splotowych. Kostka $\{-1,1\}^n$ posiada naturalną strukturę grupy abelowej, zadaną przez mnożenie

$$(x_1,\ldots,x_n)\cdot(y_1,\ldots,y_n)=(x_1y_1,\ldots,x_n,y_n).$$

Miara μ^n jest tzw. miarą Haara dla grupy $(\{-1,1\}^n, \cdot)$, mianowicie $\mu(g \cdot A) = \mu(A)$ dla dowolnego $g \in \{-1,1\}^n$ i dowolnego zbioru $A \subset \{-1,1\}^n$. Przyjęliśmy tu notację $g \cdot A = \{g \cdot a : a \in A\}.$

Dla dowolnej miary ν na zbiorze $\{-1,1\}^n$ definiujemy operator T_{ν} ,

$$T_{\nu}(f)(x) = \int f(xy^{-1}) \,\mathrm{d}\nu(y).$$

Jest to operator splotu z miarą ν . Zauważmy, że dla każdego $y \in \{-1, 1\}^n$ mamy $y^{-1} = y$, a zatem nasz operator możemy zapisać równoważnie wzorem

$$T_{\nu}(f)(x) = \int f(xy) \, \mathrm{d}\nu(y).$$

Operator T_{ν} , działający na przestrzeni $L_p(\{-1,1\}^n,\mu)$ dla $p \geq 1$, ma normę 1, czyli jest słabą kontrakcją. Faktycznie, wykorzystując fakt, że μ jest miarą Haara, otrzymujemy

$$\|T_{\nu}(f)\|_{p}^{p} = \int \left|\int f(xy^{-1}) \, \mathrm{d}\nu(y)\right|^{p} \, \mathrm{d}\mu(x) \leq \int \int \left|f(xy^{-1})\right|^{p} \, \mathrm{d}\nu(y) \, \mathrm{d}\mu(x)$$
$$= \int \int \left|f(xy^{-1})\right|^{p} \, \mathrm{d}\mu(x) \, \mathrm{d}\nu(y) = \int \int \left|f(x)\right|^{p} \, \mathrm{d}\mu(x) \, \mathrm{d}\nu(y) = \|f\|_{p}^{p}.$$

Weźmy teraz szczególną miarę ν , mianowicie

$$\nu = \nu_{\delta}^{n} = \left(\frac{1+\delta}{2}\delta_{1} + \frac{1-\delta}{2}\delta_{-1}\right)^{\otimes n}$$

i przyjmijmy oznaczenie $T_{\delta} = T_{\delta}^{(n)} = T_{\nu_{\delta}^{n}}$. Możemy bardzo łatwo stwierdzić, w jaki sposób operator T_{δ} działa na funkcjach Walsha-Fouriera,

$$T_{\delta}(w_S)(x) = \int \prod_{i \in S} x_i y_i \, \mathrm{d}\nu_{\delta}^n(y) = \left(\prod_{i \in S} x_i\right) \left(\prod_{i \in S} \int y_i \, \mathrm{d}\nu_{\delta}(y_i)\right)$$
$$= w_S(x) \delta^{|S|}.$$

Oznacza to, że funkcja w_S jest funkcją własną operatora T_{δ} i odpowiadająca jest wartość własna jest równa $\delta^{|S|}$. Jeśli zatem rozważymy dowolną funkcję $f : \{-1, 1\}^n \to \mathbb{R}$, to

$$T_{\delta}(f) = \sum_{S \subset [n]} a_s \delta^{|S|} w_S, \qquad \text{gdy} \quad f = \sum_{S \subset [n]} a_S w_S.$$

Twierdzenie 3 można zapisać zatem w równoważnej, operatorowej postaci.

Twierdzenie 4 (Bonami-Beckner-Gross). Dla każdej funkcji $f : \{-1,1\}^n \to \mathbb{R}$ i liczby $\delta \in [0,1]$ prawdziwa jest nierówność

$$||T_{\delta}f||_2 \le ||f||_{1+\delta^2}.$$

Można teraz łatwo zrozumieć, skąd bierzę się nazwa *hiperkonstrakcja*. Operator T_{δ} jest słabą kontrakcją, a zatem $||T_{\delta}f||_2 \leq ||f||_2$. Nasza nierówność jest jednak subtelniejsza, gdyż $||f||_{1+\delta^2} \leq ||f||_2$.

Następujący lemat pozwoli nam zredukować dowód powyższego twierdzenia do przypadku n = 1.

Lemat 3 (o iloczynie tensorowym operatorów spłotowych). Niech $q \ge p \ge 1$ i niech $(\Omega_1, \mu_1), (\Omega_2, \mu_2)$ będą dwoma skończonymi przestrzeniami probabilistycznymi. Niech $K_i : \Omega_i \times \Omega_i \to \mathbb{R}$ dla i = 1, 2. Definiujemy operatory

$$T_i(f)(x) = \int_{\Omega_i} K_i(x, y) \,\mathrm{d}\mu_i(y), \qquad i = 1, 2.$$

Dla funkcji $f: \Omega_1 \times \Omega_2 \to \mathbb{R}$ definiujemy iloczyn tensorowy operatorów T_1 i T_2 ,

$$(T_1 \otimes T_2)(f)(x_1, x_2) = \int_{\Omega_1} \int_{\Omega_2} f(y_1, y_2) K_1(x_1, y_1) K_2(x_2, y_2) \, \mathrm{d}\mu_2(y_2) \, \mathrm{d}\mu_1(y_1)$$

Przypuśćmy, że dla i = 1, 2 prawdziwe są nierówności

$$\|T_i f\|_{L_q(\Omega_i,\mu_i)} \le \|f\|_{L_p(\Omega_i,\mu_i)}, \qquad \text{dla wszystkich} \ f:\Omega_i \to \mathbb{R}.$$

Wówczas

$$\|T_1 \otimes T_2 f\|_{L_q(\Omega_1 \times \Omega_2, \mu_1 \otimes \mu_2)} \le \|f\|_{L_p(\Omega_1 \times \Omega_2, \mu_1 \otimes \mu_2)}$$

Dowód. Niech $f: \Omega_1 \times \Omega_2 \to \mathbb{R}$. Operator T_2 działa na funkcjach jednej zmiennej, $f: \Omega_1 \to \mathbb{R}$. Możemy jednak zdefiniować jego działanie na funkcjach dwóch zmiennych wzorem

$$T_2(f)(y_1, x_2) = \int f(y_1, y_2) K_2(x_2, y_2) \,\mathrm{d}\mu_2(y_2).$$

Zauważmy, że zgodnie z tą notacją mamy dla $f:\Omega_1\times\Omega_2\to\mathbb{R}$ relację

$$T_1 \otimes T_2 f = T_1(T_2(f)),$$

a precyzyjniej,

$$(T_1 \otimes T_2)(f)(x_1, x_2) = T_1 (T_2(f)(\cdot, x_2)) (x_1)$$

Z założenia o operatorze T_1 mamy

$$\begin{aligned} \|T_1 \otimes T_2 f\|^q_{L_q(\Omega_1 \times \Omega_2, \mu_1 \otimes \mu_2)} &= \int_{\Omega_2} \int_{\Omega_1} |T_1 \left(T_2(f)(\cdot, x_2) \right) (x_1)|^q \, \mathrm{d}\mu_1(x_1) \, \mathrm{d}\mu_2(x_2) \\ &\leq \int_{\Omega_2} \left(\int_{\Omega_1} |(T_2(f)(y_1, x_2))|^p \, \mathrm{d}\mu_1(y_1) \right)^{q/p} \, \mathrm{d}\mu_2(x_2). \end{aligned}$$

Dla dowolnych skończonych przestrzeni probabilistycznych $(X, \mu), (Y, \nu)$ i dowolnego $r \ge 1$ prawdziwa jest nierówność Minkowskiego,

$$\left(\int_X \left(\int_Y g(x,y) \,\mathrm{d}\nu(y)\right)^r \,\mathrm{d}\mu(x)\right)^{1/r} \le \int_Y \left(\int_X g(x,y)^r \,\mathrm{d}\mu(x)\right)^{1/r} \,\mathrm{d}\nu(y).$$

Faktycznie, zauważmy, że całki po przestrzeniYsą w rzeczywistości skończonymi sumami. Powyższa nierówność jest zatem nierównością postaci

$$\left\|\sum_{i} a_{i} g_{i}\right\|_{r} \leq \sum_{i} a_{i} \left\|g_{i}\right\|_{r},$$

gdzie $g_i : X \to \mathbb{R}$ i (a_i) są liczbami nieujemnymi. Zatem jest to nierówność Minkowskiego znana z kursu analizy funkcjonalnej. Skorzystamy z powyższej nierówności dla funkcji

$$g(y_1, x_2) = |(T_2(f)(y_1, x_2))|^p$$

i $(X, \mu) = (\Omega_2, \mu_2), (Y, \nu) = (\Omega_1, \mu_1) \text{ oraz } r = q/p,$
$$\left(\int_{\Omega_2} \left(\int_{\Omega_1} |(T_2(f)(y_1, x_2))|^p d\mu_1(y_1)\right)^{q/p} d\mu_2(x_2)\right)^{p/q}$$
$$\leq \left(\int_{\Omega_1} \left(\int_{\Omega_2} |(T_2(f)(y_1, x_2))|^q d\mu_2(x_2)\right)^{p/q} d\mu_1(y_1)\right).$$

Otrzymujemy

$$\int_{\Omega_2} \left(\int_{\Omega_1} |(T_2(f)(y_1, x_2))|^p \, \mathrm{d}\mu_1(y_1) \right)^{q/p} \, \mathrm{d}\mu_2(x_2) \\ \leq \left(\int_{\Omega_1} \left(\int_{\Omega_2} |(T_2(f)(y_1, x_2))|^q \, \mathrm{d}\mu_2(x_2) \right)^{p/q} \, \mathrm{d}\mu_1(y_1) \right)^{q/p}.$$

Korzystając z założenia o ${\cal T}_2$ mamy

$$\left(\int_{\Omega_2} \left| (T_2(f)(y_1, x_2)) \right|^q \, \mathrm{d}\mu_2(x_2) \right)^{1/q} \le \left(\int_{\Omega_2} \left| f(y_1, y_2) \right|^p \, \mathrm{d}\mu_2(y_2) \right)^{1/p}.$$

Stąd

$$\int_{\Omega_1} \left(\int_{\Omega_2} |(T_2(f)(y_1, x_2))|^q \, \mathrm{d}\mu_2(x_2) \right)^{p/q} \, \mathrm{d}\mu_1(y_1) \\ \leq \int_{\Omega_1} \int_{\Omega_2} |f(y_1, y_2)|^p \, \mathrm{d}\mu_2(y_2) \, \mathrm{d}\mu_1(y_1).$$

 ${\it Zatem\ ostatecznie}$

$$\|T_1 \otimes T_2 f\|_{L_q(\Omega_1 \times \Omega_2, \mu_1 \otimes \mu_2)}^q \le \left(\int_{\Omega_1} \int_{\Omega_2} |f(y_1, y_2)|^p \, \mathrm{d}\mu_2(y_2) \, \mathrm{d}\mu_1(y_1) \right)^{q/p} \\ = \|f\|_{L_p(\Omega_1 \times \Omega_2, \mu_1 \otimes \mu_2)}^q.$$

Okazuje się, że operator $T^{(n)}_\delta$ jest n-tą potęgą tensorową operatora $T^{(1)}_\delta.$ Aby się o tym przekonać zauważmy, że dla n=1mamy

$$T_{\delta}^{(1)}(f)(x) = \frac{1+\delta}{2}f(x) + \frac{1-\delta}{2}f(-x) = \int_{\{-1,1\}} f(xy)(1+\delta y) \, \mathrm{d}\mu(y)$$
$$= \int_{\{-1,1\}} f(y)(1+\delta yx^{-1}) \, \mathrm{d}\mu(y) = \int_{\{-1,1\}} f(y)K(x,y) \, \mathrm{d}\mu(y),$$

gdzie

$$K(x,y) = 1 + \delta y x^{-1}.$$

Ogólnie otrzymujemy

$$T_{\delta}^{(n)}(f)(x) = \int_{\{-1,1\}^{n}} f(x_{1}y_{1}, \dots, x_{n}y_{n}) \, \mathrm{d}\nu_{\delta}^{(1)}(y_{1}) \dots \mathrm{d}\nu_{\delta}^{(1)}(y_{n})$$

= $\int_{\{-1,1\}^{n}} f(y_{1}, \dots, y_{n})(1 + \delta y_{1}x_{1}^{-1}) \dots (1 + \delta y_{n}x_{n}^{-1}) \, \mathrm{d}\mu^{(1)}(y_{1}) \dots \mathrm{d}\mu^{(1)}(y_{n})$
= $\int_{\{-1,1\}^{n}} f(y_{1}, \dots, y_{n})K(x_{1}, y_{1}) \dots K(x_{n}, y_{n}) \, \mathrm{d}\mu^{(1)}(y_{1}) \dots \mathrm{d}\mu^{(1)}(y_{n}).$

Z powyższych rozważań i z Lematu 3 wynika, że wystarczy udowodnić Twierdzenie 4 dla $n=1.~{\rm W}$ tym przypadku mamy

$$(T_{\delta}f)(x) = \frac{1+\delta}{2}f(x) + \frac{1-\delta}{2}f(-x).$$

Zatem,

$$\|T_{\delta}f\|_{2} = \left(\frac{\left|\frac{1+\delta}{2}f(1) + \frac{1-\delta}{2}f(-1)\right|^{2} + \left|\frac{1+\delta}{2}f(-1) + \frac{1-\delta}{2}f(1)\right|^{2}}{2}\right)^{1/2}$$

oraz

$$\|f\|_{1+\delta^2} = \left(\frac{|f(1)|^{1+\delta^2} + |f(-1)|^{1+\delta^2}}{2}\right)^{\frac{1}{1+\delta^2}}.$$

Niech

$$a = \frac{f(1) + f(-1)}{2}, \qquad b = \frac{f(1) - f(-1)}{2}.$$

Nierówność $\|T_{\delta}f\|_2 \leq \|f\|_{1+\delta^2}$ jest równoważna

$$\left(\frac{|a+b\delta|^2+|a-b\delta|^2}{2}\right)^{1/2} \le \left(\frac{|a+b|^{1+\delta}+|a-b|^{1+\delta^2}}{2}\right)^{\frac{1}{1+\delta^2}}$$

Ponieważ

$$\frac{|a+b\delta|^2 + |a-b\delta|^2}{2} = a^2 + \delta^2 b^2,$$

wystarczy udowodnić następujący elementarny lemat.

Lemat 4. Dla dowolnych $a, b \in \mathbb{R}$ i $\delta \in [0, 1]$ prawdziwa jest nierówność

$$(a^{2} + b^{2}\delta^{2})^{\frac{1+\delta^{2}}{2}} \le \frac{|a+b|^{1+\delta^{2}} + |a-b|^{1+\delta^{2}}}{2}.$$

Dowód.Dlaa=0nasza nerówność ma postać $|b|^{1+\delta^2}\delta^{1+\delta^2} \leq |b|^{1+\delta^2}$, co jest prawdą, bo $\delta^{1+\delta^2} \leq 1^{1+\delta^2} = 1$. Możemy zatem zakładać, że $a \neq 0$. Dzieląc obydwie strony nierówności przez $|a|^{1+\delta^2}$ i oznaczający=b/aotrzymujemy nierówność

$$(1+\delta^2 y^2)^{\frac{1+\delta^2}{2}} \le \frac{|1+y|^{1+\delta^2}+|1-y|^{1+\delta^2}}{2}.$$

Obydwie strony nierówności są parzystą funkcją zmiennej y. Możemy zatem zakładać, że $y \geq 0.$

Rozważmy najpierw przypadek $y \in [0, 1)$. Ze wzoru Taylora mamy

$$(1+x)^{1+\delta^2} = \sum_{k=0}^{\infty} {\binom{1+\delta^2}{k}} x^k, \qquad |x| < 1,$$

gdzie

$$\binom{1+\delta^2}{k} = \frac{(1+\delta^2)(1+\delta^2-1)\dots(1+\delta^2-k+1)}{k!}$$

Stąd

$$\frac{|1+y|^{1+\delta^2} + |1-y|^{1+\delta^2}}{2} = \frac{1}{2} \left[\sum_{k=0}^{\infty} \binom{1+\delta^2}{k} y^k + \sum_{k=0}^{\infty} \binom{1+\delta^2}{k} (-y)^k \right]$$
$$= \sum_{k=0}^{\infty} \binom{1+\delta^2}{2k} y^{2k} = 1 + \frac{(1+\delta^2)\delta^2}{2} y^2 + \sum_{k=2}^{\infty} \binom{1+\delta^2}{2k} y^{2k}$$
$$\ge 1 + \frac{(1+\delta^2)\delta^2}{2} y^2,$$

ponieważ

$$\binom{1+\delta^2}{2k} = \frac{(1+\delta^2)(1+\delta^2-1)\dots(1+\delta^2-2k+1)}{(2k)!} \ge 0$$

ze względu na fakt, że w liczniku mamy 2 liczby dodatnie i 2k liczb ujemnych. Wystarczy zatem udowodnić nierówność

$$(1+\delta^2 y^2)^{\frac{1+\delta^2}{2}} \le 1 + \frac{(1+\delta^2)\delta^2}{2} y^2.$$
(6)

Zauważmy, że $(1+x)^{\lambda} \leq 1 + \lambda x$ dla $x \geq 0$ i $\lambda \in [0,1]$. Jest to wersja nierówności Bernoulliego. Wynika ona z faktu, że funkcja $g(x) = (1+x)^{\lambda} - 1 - \lambda x$ spełnia g(0) = 0 oraz $g'(x) \leq 0$ dla $x \geq 0$. Przyjmując $x = \delta^2 y^2$ oraz $\lambda = \frac{1+\delta^2}{2}$ otrzymujemy (6).

Przypadek y = 1 wynika z powyższego przez przejście graniczne.

Zajmijmy się teraz przypadkiem y>1. Niech $z=\frac{1}{y}<1.$ Musimy wykazać nierówność

$$\left(1 + \frac{\delta^2}{z^2}\right)^{\frac{1+\delta^2}{2}} \le \frac{\left|1 + \frac{1}{z}\right|^{1+\delta^2} + \left|1 - \frac{1}{z}\right|^{1+\delta^2}}{2}$$

Mnożąc obydwie strony tej nierówności przez $z^{1+\delta^2}$ otrzymujemy równoważnie

$$(z^2 + \delta^2)^{\frac{1+\delta^2}{2}} \le \frac{|1+z|^{1+\delta^2} + |1-z|^{1+\delta^2}}{2}.$$

Powyższa nierówność wynika z pierwszego przypadku, gdyż

$$z^{2} + \delta^{2} = 1 + \delta^{2} z^{2} - (1 - z^{2})(1 - \delta^{2}) \le 1 + \delta^{2} z^{2}.$$

			. 1	
		L		
		L	1	

3.2.3 Influencje funkcji boolowskich

Niech p = 1/2. Zdefiniujemy ważne pojęcie influencji funkcji boolowskiej. Z kostką dyskretną $\{-1, 1\}^n$ związana jest naturalna struktura grafowa. Mianowicie, punkty $x, y \in \{-1, 1\}^n$ są połączone krawędzią, jeśli $|\{i : x_i \neq y_i\}| = 1$. Piszemy wtedy $x \sim y$. Dla $x = (x_1, \ldots, x_i, \ldots, x_n)$ przyjmujemy oznaczenie $x^i = (x_1, \ldots, 1 - x_i, \ldots, x_n)$. Niech $f : \{-1, 1\}^n \to \{-1, 1\}$. Wyrażenie

$$I_i(f) = \mathbb{P}(f(x) \neq f(x^i)) = \frac{1}{2^n} \left| \left\{ x \in \{-1, 1\}^n : \ f(x) \neq f(x^i) \right\} \right|$$

nazywamy i-tą influencją funkcji f. Całkowitą influencją funkcji f nazywamy sumę

$$I(f) = \sum_{i=1}^{n} I_i.$$

Podamy teraz kombinatoryczną interpretację liczby I(f).Dl
a $A,B\subset\{-1,1\}^n$ definiujemy

$$E(A,B) = |\{(a,b): a \in A, b \in B, a \sim b\}|.$$

Zbiór krawędzi $E(A, A^c)$ będziemy nazywać brzegim zbioru A. Zdefiniujmy funkcję $f_A : \{-1, 1\}^n \to \{-1, 1\}$ wzorem $f_A = \mathbf{1}_A - \mathbf{1}_{A^c}$. Wówczas

$$\frac{|E(A, A^c)|}{2^{n-1}} = \frac{2|E(A, A^c)|}{2^n} = \frac{\sum_{i=1}^n |\{x : f_A(x) \neq f_A(x^i)\}|}{2^n} = \sum_{i=1}^n I_i(f_A) = I(f_A).$$

Zatem

$$I(f_A) = \frac{|E(A, A^c)|}{2^{n-1}}.$$

Pojęcie influencji możemy rozszerzyć na przypadek $p \neq 1/2$, definiując

$$I_i^p(f) = \mathbb{P}_p(f(x) \neq f(x^i)), \quad I^p(f) = \sum_{i=1}^n I_i^p(f).$$

3.2.4 Influencje vs. spektrum

Przypomnijmy, że każdą funkcję $f:\{-1,1\}^n \to \mathbb{R}$ możemy przedstawić w postaci

$$f = \sum_{S \subseteq [n]} a_s w_S,$$

gdzie $(w_S)_{s \in [n]}$ są funkcjami Walsha-Fouriera. Zauważmy, że

$$||f||_2^2 = \left\langle \sum_S a_S w_S, \sum_T a_T w_T \right\rangle = \sum_{S,T} a_S a_T \left\langle w_S, w_T \right\rangle = \sum_S a_S^2.$$

Jest tzw. tożsamość Parsevala. Niech $f_i(x)=f(x)-f(x^i).$ Łatwo zauważyć, że mamy

$$\hat{f}_i(S) = \begin{cases} 0 & i \notin S \\ 2\hat{f}(S) & i \in S \end{cases}.$$

Zatem

$$||f_i||_2^2 = 4 \sum_{S: i \in S} a_S^2.$$

Z drugiej strony,

$$|f_i(x)| = \begin{cases} 0 & f(x) = f(x^i) \\ 2 & f(x) \neq f(x^i) \end{cases}.$$

Stąd

$$||f_i||_p^p = 2^p \mathbb{P}(f(x) \neq f(x^i)) = 2^p I_i(f).$$

Przyjmując p = 2 otrzymujemy

$$I_i(f) = \sum_{S: i \in S} a_S^2,$$

a zatem

$$I(f) = \sum_{i=1}^{n} \sum_{S: i \in S} a_{S}^{2} = \sum_{S} |S| a_{S}^{2}.$$

Tożsamość ta wiąże influencję funkcji boolowskiej z jej spektrum.

Zdefiniujmy

$$\operatorname{Var}_{\mu}(f) = \mathbb{E}_{\mu}f^2 - (\mathbb{E}_{\mu}f)^2.$$

Zauważmy, że

$$\mathbb{E}f = \sum_{S} a_{S} \mathbb{E}w_{S} = a_{\emptyset}.$$

Stąd

$$\operatorname{Var}_{\mu}(f) = \sum_{S} a_{S}^{2} - a_{\emptyset}^{2} = \sum_{S: |S| \ge 1} a_{S}^{2}.$$

Z drugiej strony, przyjmując $\mu(f) = \mathbb{P}(f = 1)$, otrzymujemy

$$\operatorname{Var}_{\mu}(f) = \mathbb{E}_{\mu}f^{2} - \left(\mathbb{E}_{\mu}f\right)^{2} = 1 - \left(\mathbb{P}(f=1) - \mathbb{P}(f=-1)\right)^{2}$$
$$= 1 - \left(2\mu(f) - 1\right)^{2} = 4\mu(f)(1 - \mu(f)).$$

Możemy teraz wykorzystać wprowadzone pojęcia w celu wykazaia, że funkcja $\text{Dict}_n(x_1, \ldots, x_n) = x_1$ ma najmniejszą influencję spośród wszystkich funkcji o średniej 0. Innymi słowy, brzeg każdego podzbioru kostki mającego 2^{n-1} elementów składa się z conajmniej 2^{n-1} krawędzi.

Twierdzenie 5. Niech $f : \{-1, 1\}^n \to \{-1, 1\}$ i niech $\mu(f) = \mathbb{P}(f = 1)$. Wówczas mamy

$$I(f) \ge 4\mu(f)(1-\mu(f)).$$

W szczególności dla $\mu(f) = 1/2$ otrzymujemy $I(f) \ge 1$.

Dowód. Mamy

$$4\mu(f)(1-\mu(f)) = \operatorname{Var}_{\mu}(f) = \sum_{S: |S| \ge 1} a_{S}^{2} \le \sum_{S: |S| \ge 1} |S|a_{S}^{2} = \sum_{S} |S|a_{S}^{2} = I(f).$$

3.2.5 Twierdzenia KKL i Twierdzenia Talagranda

Możemy teraz sformułować główne twierdzenia tego rozdziału.

Twierdzenie 6 (Kahn-Kalai-Linial, [KKL]). Niech p = 1/2 i niech $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$. Wówczas

$$\max_{1 \le i \le n} I_i(f) \ge \frac{2}{15} \operatorname{Var}(f) \frac{\ln n}{n}.$$

Powyższe twierdzenie można rozszerzyć na przypadek $p \neq 1/2$.

Twierdzenie 7 (Bourgain, Kahn, Kalai, Katznelson, Linial, [BKKKL]). Niech $f : \{-1, 1\}^n \to \{-1, 1\}$ będzie funkcją monotoniczną. Wówczas

$$\max_{1 \le i \le n} I_i^p(f) \ge \frac{2}{15} \operatorname{Var}_p(f) \frac{\ln n}{n}.$$

Prawdziwe są również wzmocnione wersje tych oszacować, patrz [T]. Przyjmujemy tutaj notację $\frac{0}{\log(1/0)} = 0$ oraz $1/\log(1) = +\infty$.

Twierdzenie 8 (Talagrand). Niech p = 1/2 i niech $f : \{-1, 1\}^n \to \{-1, 1\}$. Wówczas

$$I(f) \ge \frac{4}{15} \operatorname{Var}(f) \ln \left(\frac{1}{\max_{1 \le i \le n} I_i(f)} \right).$$

Twierdzenie 9 (Talagrand). Niech p = 1/2 i niech $f : \{-1, 1\}^n \to \{-1, 1\}$. Wówczas

$$\sum_{i=1}^{n} \frac{I_i(f)}{\ln(1/I_i(f))} \ge \frac{4}{15} \operatorname{Var}(f).$$

Twierdzenie 10 (Talagrand). Niech $f:\{-1,1\}^n\to\{-1,1\}$ będzie funkcją monotoniczną. Wówczas

$$I^p(f) \ge c \operatorname{Var}_p(f) \ln\left(\frac{1}{\max_{1 \le i \le n} I^p_i(f)}\right)$$

Nasza strategia dowodu wymienionych twierdzeń jest następująca. Najpierw udowodnimy Twierdzenie 9. Następnie, korzystając z tego twierdzenia wykażenie Twierdzenie 8. Potem wywnioskujemy Twierdzenie 10. Następnie udowodnimy Twierdzenia 6 i 7. Naszym najważniejszym narzędziem będzie hiperkontrakcja.

Będzie nam potrzebny następujący lemat.

Lemat 5. Niech $g: \{-1,1\}^n n \to \mathbb{R}$ oraz $||g||_{3/2} \neq ||g||_2$ (równoważnie |g| nie jest stała). Wówczas

$$\sum_{S \neq \emptyset} \frac{\hat{g}(S)^2}{|S|} \le \frac{5}{2} \frac{\|g\|_2^2}{\log\left(\|g\|_2 / \|g\|_{3/2}\right)}.$$

Dowód. Skorzystajmy z hiperkostrakcji

$$\|T_{\delta}g\|_{2} \leq \|g\|_{1+\delta^{2}},$$

przyjmując $\delta^2=1/2.$ Otrzymujemy

$$\sum_{S: |S|=k} \hat{g}(S)^2 \le 2^k \sum_{S} \frac{1}{2^{|S|}} \hat{g}(S)^2 = 2^k \left\| T_{\sqrt{1/2}} g \right\|_2^2 \le 2^k \left\| g \right\|_{3/2}^2.$$

Niech $m \ge 0$. Mamy

$$\sum_{S \neq \emptyset} \frac{\hat{g}(S)^2}{|S|} = \sum_{k=1}^m \sum_{S: \ |S|=k} \frac{\hat{g}(S)^2}{k} + \sum_{S: \ |S|>m} \frac{\hat{g}(S)^2}{|S|} \le \sum_{k=1}^m \frac{2^k \|g\|_{3/2}^2}{k} + \sum_{S: \ |S|>m} \frac{\hat{g}(S)^2}{m+1} \le \frac{4 \cdot 2^m \|g\|_{3/2}^2 + \|g\|_2^2}{m+1},$$

gdzie skorzystaliśmy z elementarnej nierówności

$$\sum_{k=1}^m \frac{2^k}{k} \le \frac{4 \cdot 2^m}{m+1},$$

którą można wykazać indukcyjnie.

Przyjmijmy

$$m = \max\{m \ge 0 \mid 2^m \|g\|_{3/2}^2 \le \|g\|_2^2\}.$$

Wtedy $2^{m+1} \|g\|_{3/2}^2 > \|g\|_2^2$. Zatem,

$$m+1 > 2\log\left(\frac{\|g\|_2}{\|g\|_{3/2}}\right)$$

Otrzymujemy

$$\sum_{S \neq \emptyset} \frac{\hat{g}(S)^2}{|S|} \le \frac{5 \|g\|_2^2}{m+1} \le \frac{5}{2} \frac{\|g\|_2^2}{\log\left(\|g\|_2 / \|g\|_{3/2}\right)}.$$

L		
L		
L		

Dowód Twierdzenia 9. Niech $g(x) = f(x) - f(x^i)$. Załóżmy, że $I_i(f) \in (0, 1)$. Zatem funkcja |g| nie jest stała. Mamy

$$\frac{\|g\|_2}{\|g\|_{3/2}} = \frac{2I_i(f)^{1/2}}{2I_i(f)^{2/3}} = I_i(f)^{-1/6}.$$

Korzystając z lematu mamy

$$\sum_{S:\ i\in S} \frac{4\hat{f}(S)^2}{|S|} = \sum_{S} \frac{\hat{g}(S)^2}{|S|} \le \frac{5}{2} \frac{\|g\|_2^2}{\log\left(\|g\|_2 / \|g\|_{3/2}\right)} = \frac{5}{2} \cdot \frac{4I_i(f)}{\log(I_i(f)^{-1/6})} = 60 \frac{I_i(f)}{\log(\frac{1}{I_i(f)})}.$$

Nierówność

$$\sum_{S:\ i \in S} \frac{4\hat{f}(S)^2}{|S|} \le 60 \frac{I_i(f)}{\log(\frac{1}{I_i(f)})}$$

jest również prawdziwa w przypadku $I_i(f) \in \{0, 1\}$. Otrzymujemy

$$16\mu(f)(1-\mu(f)) = 4\operatorname{Var}_{\mu}(f) = \sum_{Sn\in\emptyset} 4\hat{f}(S)^2 = \sum_{i=1}^n \sum_{S:\ i\in S} \frac{4\hat{f}(S)^2}{|S|} \le 60\sum_{i=1}^n \frac{I_i(f)}{\log(\frac{1}{I_i(f)})}.$$

Wynika stąd już dowodzona nierówność.

Pokażemy teraz, że Twierdzenie 9 implikuje Twierdzenie 8.

Dowód Twierdzenia 8. Wystarczy zauważyć, że

$$\frac{I(f)}{\ln\left(1/\max_{1\le i\le n}I_i(f)\right)} \ge \frac{\sum_{i=1}^n I_i(f)}{\ln\left(1/\max_{1\le i\le n}I_i(f)\right)} \ge \sum_{i=1}^n \frac{I_i(f)}{\ln\left(1/I_i(f)\right)} \ge \frac{4}{15} \operatorname{Var}(f).$$

Pokażemy, w jaki sposób z Twierdzenia 8 można uzyskać Twierdzenie 10.

Dowód Twierdzenia 10. Po pierwsze zauważmy, że Twierdzenie 10 wystarczy wykazać dla $p = \frac{k}{2^n}$, gdzie $n \ge 1$ i $0 \le k \le 2^n$. Wynika to z ciągłości obydwu stron nierówności względem parametru p. Rozpatrzmy przekształcenie kodujące (a w zasadzie dekodujące) $\pi : \{0, 1\}^l \to \{0, 1/2^l, \dots, (2^l - 1)/2^l\}$ zadane wzorem

$$\pi(x_1,\ldots,x_l) = \sum_{i=1}^l x_i/2^i.$$

To przekształcenie jest bijekcją, a zatem jeśli x ma rozkład jednostajny na $\{0, 1\}^l$, to $\pi(x)$ ma rozkład jednostajny na $\{0, 1/2^l, \ldots, (2^l - 1)/2^l\}$, a zatem $\mathbb{P}\left(\pi(x \ge i/2^l)\right) = (2^l - i)/2^l$. Definiujemy przekształcenie $g: \{0, 1\}^l \to \{0, 1\}$ zadane wzorem $g(x) = \mathbf{1}_{\{\pi(x) \ge 1-p\}}$. Oczywiście

$$\mathbb{P}\left(g(x)=1\right) = \mathbb{P}\left(\pi(x) \ge 1-p\right) = p.$$

Definiujemy $\tilde{f}: \{0,1\}^{nl} \to \{0,1\}$ wzorem

$$\tilde{f}(x_1^1, \dots, x_l^1, x_1^2, \dots, x_l^2, \dots, x_1^n, \dots, x_l^n) = f(g(x_1^1, \dots, x_l^1), g(x_1^2, \dots, x_l^2), \dots, g(x_1^n, \dots, x_l^n)).$$

Z powyższych uwag wynika natychmiast, że \tilde{f} , jako funkcja zdefiniowana na przestrzeni probabilistycznej $\left(\{0,1\}^{nl}, \left(\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1\right)^{nl}\right)$ ma taki sam rozkład jak funkcja fzdefiniowana na przestrzeni probabilistycznej $(\{0,1\}^n, (p\delta_0 + (1-p)\delta_1)^n)$. W szczególności $\operatorname{Var}_p(f) = \operatorname{Var}_{1/2}(\tilde{f})$.

Udowodnimy nierówność

$$I_{(r,j)}(\tilde{f}) \le I_r^p(f)/2^{j-1}, \quad r = 1, \dots, n, \quad j = 1, \dots, l.$$
 (7)

Możemy zakładać, że r = 1. Niech $x^r = (x_1^r, \ldots, x_n^r), r = 1, \ldots, n$. Zauważmy, że

$$\begin{split} I_{(1,j)}(\tilde{f}) &= \mathbb{P}\big(\tilde{f}(x_1^1, ..., x_j^1, ..., x_n^1, ...) \neq \tilde{f}(x_1^1, ..., 1 - x_j^1, ..., x_n^1, ...)\big) \\ &= \mathbb{P}\big(g(x_1^1, ..., x_j^1, ..., x_l^1) \neq g(x_1^1, ..., 1 - x_j^1, ..., x_l^1) \text{ oraz} \\ &\quad f(0, g(x^2), ..., g(x^n)) \neq f(1, g(x^2), ..., g(x^n))\big). \end{split}$$

Niech

$$A = \{g(x_1^1, ..., x_j^1, ..., x_l^1) \neq g(x_1^1, ..., 1 - x_j^1, ..., x_l^1)\}$$

oraz

$$B = \{f(0, g(x^2), ..., g(x^n)) \neq f(1, g(x^2), ..., g(x^n))\}$$

Zdarzenia A i B są niezależne. Zatem $I_{(1,j)}(\tilde{f}) = \mathbb{P}(A) \mathbb{P}(B)$. Zauważmy, że zmienne $g(x^2), \ldots, g(x^n)$ są niezależne i mają rozkład $p\delta_0 + (1-p)\delta_1$. Zatem $\mathbb{P}(B) = I_r^p(f)$. Pozostaje wykazać, że $\mathbb{P}(A) \leq 1/2^{j-1}$. Zauważmy, że

$$\left|\pi(x_1^1,...,x_j^1,...,x_l^1) - \pi(x_1^1,...,1-x_j^1,...,x_l^1)\right| = 1/2^j.$$

Stąd

$$\mathbb{P}(A) = \mathbb{P}\left(g(x_1^1, ..., x_j^1, ..., x_l^1) \neq g(x_1^1, ..., 1 - x_j^1, ..., x_l^1)\right)$$

$$\leq \mathbb{P}\left(\pi(x) \in [1 - p - 1/2^j, 1 - p + 1/2^j)\right) = 1/2^{j-1}.$$

	_	_	_	

Twierdzenie 6 wynika z Twierdzenia 8 w identyczny sposób, w jaki Twierdzenie 7 wynika z Twierdzenia 10. Udowodnimy Twierdzenie 7.

Twierdzenie 10 implikuje Twierdzenie 7. Zauważmy, że jeśli $a \in (0, 1)$ oraz $\frac{a}{\log(1/a)} \ge c > 0$ to $a \ge \frac{1}{2}c\log(1/c)$. Ponieważ $(0, 1) \ni a \mapsto \frac{a}{\log(1/a)}$ jest rosnąca, wystarczy zakładać, że $\frac{a}{\log(1/a)} = c$. Musimy zatem wykazać nierówność

$$a \ge \frac{1}{2} \frac{a}{\log(1/a)} \log\left(\frac{1}{a}\log\left(\frac{1}{a}\right)\right).$$

Biorąc $x = 1/a \ge 1$ widzimy, że nierówność ta jest równoważna z

$$\log(x) \ge \frac{1}{2}\log(x\log(x)) = \frac{1}{2}\log x + \frac{1}{2}\log\log x$$

Wystarczy zatem wykazać $x \ge \log x$. Wynika to z nierówności Bernoulliego,

$$2^x = (1+1)^x \ge 1 + x \ge x.$$

Niech $\|I^p\|_{\infty} = \max_{1 \leq in} I^p_i(f)$. Z Twierdzenia 10 mamy

$$n \left\| I^p \right\|_{\infty} \ge I^p(f) \ge \frac{4}{15} \operatorname{Var}_p(f) \ln \left(\frac{1}{\left\| I^p \right\|_{\infty}} \right).$$

Przyjmując $a = \|I^p\|_{\infty}$ i $c = \frac{4\operatorname{Var}_p(f)}{15n}$ otrzymujemy

$$||I^p||_{\infty} \ge \frac{2\operatorname{Var}_p(f)}{15n} \ln\left(\frac{15n}{4\operatorname{Var}_p(f)}\right) \ge \frac{2\operatorname{Var}_p(f)}{15n} \ln n.$$

Skorzystaliśmy z faktu, że $\operatorname{Var}_p(f) \leq 1$.

3.3 Lemat Russo

W tym podrozdziale sformułujemy i udowodnimy słynny Lemat Margulisa-Russo. Pozwoli on nam kontrolować funkcję $\theta_f(p)$ za pomocą wielkości $I^p(f)$.

Lemat 6 (Margulis-Russo lemma). Niech $f:\{-1,1\}^n\to\{-1,1\}$ będzie funkcją monotoniczną. Wtedy

$$\frac{\mathrm{d}}{\mathrm{d}p}\mu_p(f) = I^p(f).$$

 -	_	_	

Dowód. Zdefiniujmy

$$\mu_{p_1,\dots,p_n} = \left((1-p_1)\delta_{\{-1\}} + p_1\delta_{\{1\}} \right) \otimes \dots \otimes \left((1-p_n)\delta_{\{-1\}} + p_n\delta_{\{1\}} \right).$$

Udowodnimy, że

$$\frac{\partial \mu_{p_1,\dots,p_n}(f)}{\partial p_i} = I_i^{(p_1,\dots,p_n)}(f).$$

Wówczas otrzymamy

$$\frac{\mathrm{d}\mu_p(f)}{\mathrm{d}p} = \sum_{i=1}^n \frac{\partial\mu_{p_1,\dots,p_n}(f)}{\partial p_i}\Big|_{p_1=\dots=p_n=p} = \sum_{i=1}^n I_i^{(p,\dots,p)}(f) = \sum_{i=1}^n I_i^p(f).$$

Bez zmniejszenia ogólności możemy założyć, ż
ei=1.Niech $f_1(x)=f(x)-f(x^1).$ Mamy

$$\mathbb{P}_{p_1,\dots,p_n}(f=1) = \mathbb{P}_{p_1,\dots,p_n}(f=1, f_1 \neq 0) + \mathbb{P}_{p_1,\dots,p_n}(f=1, f_1=0).$$

Definiujemy zbiór $A \subset \{-1,1\}^{n-1},$

$$A = \left\{ x \in \{-1, 1\}^{n-1} \mid f(1, x) = 1, f_1(1, x) = 0 \right\}.$$

Jeśli f(1,x) = 1 i $f_1(1,x) = 0$ to f(-1,x) = 1 i $f_1(-1,x) = 0$. Zatem

$${f = 0, f_i = 0} = {-1, 1} \times A.$$

Stąd

$$\mathbb{P}_{p_1,\ldots,p_n}(f=1,f_1=0)=\mathbb{P}_{p_2,\ldots,p_n}(A),$$

a zatem jest to wyrażenie niezależne od p_1 .

Ze względu na to, że fjest monotoniczna, otrzymujemy

$${f = 1, f_1 \neq 0} = {(x_1, \dots, x_n) | x_1 = 1, f(1, \dots, x_n) = 1, f(-1, \dots, x_n) = -1, }.$$

Defini
ujemy $B \subset \{-1,1\}^{n-1},$

$$B = \left\{ x \in \{-1, 1\}^{n-1} \mid f(1, x) = 1, f_1(1, x) \neq 0 \right\}.$$

Mamy

$$\{f = 1, f_1 = 0\} = \{1\} \times B.$$

Stąd,

$$\mathbb{P}_{p_1,...,p_n}(f = 1, f_1 \neq 0) = p_1 \mathbb{P}_{p_2,...,p_n}(B)$$

Zauważmy również, że

$$I_1^{(p_1,\dots,p_n)}(f) = \mu_{p_1,\dots,p_n}\left(\{-1,1\} \times B\right) = \mathbb{P}_{p_2,\dots,p_n}(B).$$

Dlatego

$$\frac{\partial \mu_{p_1,\dots,p_n}(f)}{\partial p_1} = \frac{\partial}{\partial p_1} \left(\mathbb{P}_{p_2,\dots,p_n}(A) + p_1 \mathbb{P}_{p_2,\dots,p_n}(B) \right) = \mathbb{P}_{p_2,\dots,p_n}(B) = I_1^{(p_1,\dots,p_n)}(f).$$

3.4 Dowód Twierdzenia Friedguta-Kalaia

Dowód.Zauważmy, że $\mu_p(f)=\theta_f(p).$ Z symetrii funkcji fmamy $I_1^p(f)=\ldots=I_n^p(f).$ Z Twierdzenia 7 i z Lematu Margulisa-Russo otrzymujemy

$$\frac{\mathrm{d}\theta(p)}{\mathrm{d}p} = I^p(f) = n \max_{1 \le i \le n} I^p_i(f) \ge \frac{8}{15} \theta(p)(1 - \theta(p)) \ln n \ge \frac{4}{15} \min\{\theta(p), 1 - \theta(p)\} \ln n.$$

Niech p_{\star} spełnia $\theta(p_{\star}) = 1/2$. Oczywiście $p_{-}(\varepsilon) < p_{\star} < p_{+}(\varepsilon)$. Dla $p \in [p_{-}(\varepsilon), p_{\star}]$ mamy $\frac{\mathrm{d}\theta(p)}{\mathrm{d}p} \geq \frac{4}{15}\theta(p)\ln n$, czyli $\frac{\mathrm{d}\ln\theta(p)}{\mathrm{d}p} \geq \frac{4}{15}\ln n$. Wynika stąd

$$\ln(1/2\varepsilon) = \ln \theta(p_{\star}) - \ln \theta(p_{-}(\varepsilon)) \ge \frac{4}{15}(p_{\star} - p_{-}(\varepsilon)) \ln n.$$

Podobnie dla $p \in [p_{\star}, p_+]$ mamy $-\frac{\mathrm{d}}{\mathrm{d}p} \ln(1 - \theta(p)) \ge \frac{4}{15} \ln n$, czyli

$$\ln(1/2\varepsilon) = -\left(\ln(1-\theta(p_{\star}(\varepsilon))) - \ln(1-\theta(p_{\star}))\right) \ge \frac{4}{15}(p_{\star}(\varepsilon) - p_{\star})\ln n.$$

Zatem

$$p_+(\varepsilon) - p_-(\varepsilon) \le \frac{15}{2} \ln(1/2\varepsilon) \frac{1}{\ln n}$$

-		
L		
L		



Rysunek 10: Krata \mathbb{Z}^2 .

4 Perkolacje na kracie \mathbb{Z}^2

4.1 Prawdopodobieństwo krytyczne i nierówność FKG

W tym rozdziale zajmiemy się perkolacją na kracie całkowitoliczbowej \mathbb{Z}^2 . Wprowadzamy na tej kracie strukturę grafu – punkty $x, y \in \mathbb{Z}^2$ są połączone krawędzią jeśli |x - y| = 1 (Rysunek 10).

Wprowadzamy następujący model grafu losowego. Każda krawędź (nieskończonej!) kraty \mathbb{Z}^2 jest otwarta z prawdopodobieństwem p i zamknięta z prawdopodobieństwem 1-p (w modelu Erdösa-Renyi pisaliśmy jest z prawdopodobieństwe p i nie ma jej z prawdopodobieństwem 1-p). Przestrzenią probabilistyczną dla tego grafu losowego jest zbiór $\{O, Z\}^{\mathbb{Z}^2}$ (gdzie symbol O oznacza krawędź otwartą, zaś symbol Z krawędź zamkniętą) wraz z σ -ciałem produktowym i miarą produktową.

Otwartym klastrem nazywamy spójną składową złożoną z punktów kraty \mathbb{Z}^2 , połączonych otwartymi krawędziami. Innymi słowy, dwa punkty leżą w jednej spójnej składowej, jeśli da się przejść od jednego punktu do drugiego, poruszając się jedynie po otwartych krawędziach. Niech E_{∞} oznacza zdarzenie, że w opisanym grafie losowym istnieje nieskończony spójny klaster. Zauważmy, że jeśli dla pewnej konfiguracji (czyli wyboru otwartych i zamkniętych krawędzi) E_{∞} zachodzi, to po zmianie skończonej liczby krawędzi z otwartych na zamknięte (lub na odwrót) zdarzenie E_{∞} nadal zachodzi. Podobnie, jeśli zdarzenie E_{∞} nie zachodzi to nic się z tym nie da zrobić, zmieniając jedynie skończenie wiele krawędzi. Wynika stąd, że E_{∞} należy do σ -ciała resztkowego, a zatem z prawa 0–1 Kołomogorowa $\mathbb{P}_p(E_{\infty}) \in \{0,1\}$. Okazuje się, że $\mathbb{P}_p(E_{\infty}) = 0$ dla $p \leq 1/2$ oraz $\mathbb{P}_p(E_{\infty}) = 1$ dla p > 1/2. Jest to słynne twier-



Rysunek 11: Przykładowa konfiguracja krawędzi w modelu perkolacji na kracie \mathbb{Z}^2 .

dzenie Harrisa-Kestena. Pierwsza część sformułowania została udowodniona w 1960 roku przez Harrisa. Druga część jest znacznie trudniejsza i została udowodniona w 1980 roku przez Kestena. Celem tego rozdziału jest przedstawienie prostego dowodu twierdzenia Harrisa-Kestena, pochodzącego od Bollobása i Riordana, [BR].

Niech $v \in \mathbb{Z}^2$ i niech C_v oznacza otwarty klaster zawierający punkt v. Niech $\theta(p) = \mathbb{P}_p(|C_0| = \infty)$, gdzie 0 oznacza punkt $(0,0) \in \mathbb{Z}^2$. Funkcja $\theta(p)$ jest niemalejącą funkcją $p \in [0,1]$. Zauważmy, że z translacyjnej niezmienniczości mamy $\mathbb{P}_p(|C_0| = \infty) = \theta(p)$ dla każdego wierzchołka $v \in \mathbb{Z}^2$. Oczywiście

$$E_{\infty} = \bigcup_{v \in \mathbb{Z}^2} \{ |C_v| = \infty \}.$$

Zatem jeśli $\theta(p) = 0$, to $\mathbb{P}_p(E_{\infty}) = 0$ oraz jeśli $\theta(p) > 0$, to $\mathbb{P}_p(E_{\infty}) = 1$.

Definiujemy prawdopodobieństwo krytyczne,

$$p_{c} = \inf\{p : \theta(p) > 0\} = \inf\{p : \mathbb{P}_{p}(E_{\infty}) > 0\} = \inf\{p : \mathbb{P}_{p}(E_{\infty}) = 1\}.$$

W dowodzie twierdzenia skorzystamy z Twierdzenia 1.

Niech X będzie zbiorem N-elementowym. Niech X_p będzie losowym podzbiorem zbioru X, przy czym, niezależnie, każdy element jest w zbiorze X_p z prawdopodobieństwem p. Rodzina $\mathcal{A} \subseteq \mathcal{P}(X)$ jest rosnąca, jeśli $A \in \mathcal{A}$ oraz $A \subseteq B \subset X$ implikują $B \in \mathcal{A}$. Rodzina \mathcal{A} jest malejąca, jeśli $\mathcal{P}(X) \setminus \mathcal{A}$ jest rosnąca. Będzie nam potrzebny następujący lemat.

Lemat 7. (Nierówność FKG) Niech \mathcal{A}, \mathcal{B} będą dwiema rosnącymi lub dwiema malejącymi rodzinami podzbiorów zbioru X. Wówczas w modelu losowego podzbioru X_p mamy

$$\mathbb{P}_{p}\left(\mathcal{A}\cap\mathcal{B}\right)\geq\mathbb{P}_{p}\left(\mathcal{A}\right)\mathbb{P}_{p}\left(\mathcal{B}\right).$$

Dowód. Udowodnimy więcej. Wprowadzimy pojęcie rosnącej zmiennej losowej. Zmienna losowa S określona na $\mathcal{P}(X)$ jest rosnąca jeśli $A \subseteq B \subseteq X$ implikuje $S(A) \leq S(B)$. Oczywiście rodzina \mathcal{A} jest rosnąca wtedy i tylko wtedy gdy $\mathbf{1}_{\mathcal{A}}$ jest rosnącą zmienną losową. Ponieważ $\mathbb{P}_p(\mathcal{A}) = \mathbb{E} \mathbf{1}_{\mathcal{A}}$ nasza nierówność jest równoważna z

$$\mathbb{E}_p ST \ge \mathbb{E}_p S\mathbb{E}_p T,\tag{8}$$

gdzie $S = \mathbf{1}_{\mathcal{A}}$ oraz $T = \mathbf{1}_{\mathcal{B}}$. Wykażemy (8) dla dowolnych rosnących zmiennych losowych X, Y. Bez straty ogólności możemy przyjąć, że $X = \{1, 2, ..., N\}$. Przeprowadzimy dowód indukcyjny ze względu na N. Podzbiór $A \subseteq X$ będziemy identyfikowali z N-elementowym ciągiem binarnym, przy czym 0 na *i*-tym miejscu oznacza, że elementu *i* nie ma w zbiorze A. Np., dla N = 4 ciąg (1, 1, 0, 1) oznacza podzbiór $\{1, 2, 4\}$. Będziemy też pisali X(1, 1, 0, 1) zamiast $X(\{1, 2, 4\})$.

(N = 1) Mamy

$$\mathbb{E}_p ST - \mathbb{E}_p S\mathbb{E}_p T = S(1)T(1)p + S(0)T(0)(1-p) - S(1)T(1)p^2 - S(0)T(0)(1-p)^2 - S(0)T(1)p(1-p) - S(1)T(0)p(1-p) = p(1-p)(S(1) - S(0))(T(1) - T(0)) \ge 0,$$

bo $S(1) \ge S(0)$ i $T(1) \ge T(0)$, ze względu na monotoniczność T i S. (Krok indukcyjny $n \implies n+1$) Definiujemy

$$S_1 = pS(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n, 1) + (1-p)S(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n, 0)$$

$$T_1 = pT(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n, 1) + (1-p)T(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n, 0)$$

Zmienne S_1 i T_1 , jako kombinacje zmiennych monotonicznych na $\mathcal{P}(\{1, 2, \ldots, n, n + 1\})$ są monotoniczne na $\mathcal{P}(\{1, 2, \ldots, n\})$. Stąd

$$\mathbb{E}_p S_1 T_1 \ge \mathbb{E}_p S_1 \mathbb{E}_p T_1 = \mathbb{E}_p S \mathbb{E}_p T$$

Wystarczy teraz pokazać nierówność $\mathbb{E}_p ST \geq \mathbb{E}_p S_1 T_1$. Mamy

$$\begin{split} \mathbb{E}_p ST - \mathbb{E}_p S_1 T_1 &= p \mathbb{E}_p S(\cdot, 1) T(\cdot, 1) + (1 - p) \mathbb{E}_p S(\cdot, 0) T(\cdot, 0) \\ &- \mathbb{E}_p \left(\left(p S(\cdot, 1) + (1 - p) S(\cdot, 0) \right) \left(p T(\cdot, 1) + (1 - p) T(\cdot, 0) \right) \right) \\ &= p (1 - p) \mathbb{E}_p \left(S(\cdot, 1) - S(\cdot, 0) \right) \left(T(\cdot, 1) - T(\cdot, 0) \right) \ge 0. \end{split}$$



Rysunek 12: Krata \mathbb{Z}^2 i krata dualna $\mathbb{Z}^2 + (1/2, 1/2)$.

Nierówność dla rodzin malejących udowadniamy biorąc rosnące zmienne losowe $S = -\mathbf{1}_{\mathcal{A}}$ i $T = -\mathbf{1}_{\mathcal{B}}$ i korzystając z (8).

4.2 Krata dualna

Kratą dualną do kraty \mathbb{Z}^2 nazywamy graf, którego wierzchołkami są środki ścian kraty \mathbb{Z}^2 , czyli punkty należące do zbioru $\mathbb{Z}^2 + (1/2, 1/2)$, zaś krawędzie poprowadzone są między wierzchołkami odległymi o 1 w odległości euklidesowej⁷. Oczywiście krata dualna do kraty \mathbb{Z}^2 jest z tą kratą izomorficzna (Rysunek 12).

Zauważmy, że każdą krawędź kraty dualnej przecina dokładnie jedna krawędź kraty \mathbb{Z}^2 . Zatem dla dowolnej konfiguracji krawędzi na kracie \mathbb{Z}^2 możemy zdefiniować dualną do niej konfigurację krawędzi na kracie dualnej – krawędz w kracie dualnej jest otwarta wtedy i tylko wtedy, gdy przecinająca ją krawędź jest zamknięta (Rysunek 13).

Niech a, b, c, d będą liczbami całkowitymi, przy czym a < b i c < d. Zbiór postaci $R = ([a, b] \times [c, d]) \cap \mathbb{Z}^2$ nazywamy prostokątem. Jeśli k = b - a i l = d - c, to prostokat R ma szerokość k i wysokość l. Powiemy wówczas, że R jest prostokąte $k \times l$. Zbiór $\{a\} \times \{c, \ldots, d\}$ nazywamy lewym bokiem prostokąta R. Analogicznie definiujemy prawy, górny i dolny bok.

⁷Podobnie można zdefiniować graf dualny do dowolnego grafu planarnego.



Rysunek 13: Przykładowa konfiguracja krawędzi na kracie \mathbb{Z}^2 (linie ciągłe) i odpowiadająca jej konfiguracja dualna na kracie $\mathbb{Z}^2 + (1/2, 1/2)$ (linie przerywane).



Rysunek 14: Przykład ścieżki łączącej lewy i prawy bok prostokąta.

Dla prostokąta R definiujemy prostokąt dualny

$$R^{d} = ([a+1/2, b-1/2] \times [c-1/2, d+1/2]) \cap (\mathbb{Z}^{2} + (1/2, 1/2)).$$

Prostokąt dualny do prostokąta $k \times l$ jest prostokątem $(k-1) \times (l+1)$.

Dla prostokąta R niech H(R) będzie zdarzeniem polegającym na istnieniu ścieżki złożonej z otwartych krawędzi, łączącej lewy i prawy bok (Rysunek 14). Analogicznie V(R) jest zdarzeniem polegającym na istnieniu ścieżki złożonej z otwartych krawędzi, łączącej górny i dolny bok.

Udowodnimy intuicyjnie oczywisty (ale nietrywialny w dowodzie!) lemat mówiący, że dla ustalonego prostokąta R albo istnieje otwarta ścieżka łącząca lewy i prawy bok w R albo istnieje otwarta ścieżka łącząca górny i dolny bok w R^d .

Lemat 8. Dla ustalonego prostokąta R i dla ustalonej konfiguracji krawędzi oraz indukowanej przez nią konfiguracji krawędzi na kracie dualnej zachodzi dokładnie jedno ze zdarzeń H(R), $V(R^d)$.

Dowód. Wprowadzamy pomocniczy prostokąt $([a+1/4, b-1/4] \times [c-1/4, d+1/4]) \cap (\frac{1}{2}\mathbb{Z}^2 + (1/4, 1/4))$ (patrz Rysunek 15). Bez zmniejszenia ogólności możemy przyjąć, że sąsiadujące ze sobą punkty z lewego i prawego boku prostokąta R są połączone otwartymi krawędziami. Pododnie przyjmujemy, że otwartymi krawędziami połączone sa sąsiadujące ze sobą punkty górnego i dolnego boku R^d . Między sąsiednimi wierzchołkami pomocniczego prostokąta prowadzimy zorientowaną krawędź, jeśli krawędź ta nie przetnie żadnej otwartej krawędzi w R i R^d . Ponadto poprowadzona krawędź jest skierowana tak, aby po lewej jej stronie w odległości 1/4 znajdował się



Rysunek 15: Fragment kraty \mathbb{Z}^2 i kraty dualnej. Prostokąt R na kracie \mathbb{Z}^2 (wierzchołki to małe wypełnione kropki), prostokąt R^d (duże wypełnione kropki) i pomocniczy prostokąt (białe kropki).

punkt z R lub po prawej stonie w odległości 1/4 znajdował się punkt z R^d . Przykładową konfigurację pokazano na Rysunku 16.

Łatwo zauważyć, że każdy punkt pomocniczego prostokąta ma stopień 2, poza punktami narożnymi x, y, w, z. Ponieważ stopień wierzchołka zależy od konfiguracji w jego bliskim otoczeniu, prawdziwość powyższej uwagi można sprawdzić analizując kilka prostych przypadków. Zauważmy, że wierzchołki stopnia 1, z których wychodzi krawędź to wierzchołki x oraz z. Wierzchołki stopnia 1, do których krawędź wchodzi to wierzchołki y oraz w. Ścieżka rozpoczynająca się w x musi zatem skończyć się w punkcie y lub w. Tak samo rzecz się ma dla ścieżki wychodzącej z punktu z. Ostatecznie nasz pomocniczy graf jest sumą rozłącznych cykli i dwóch ścieżek zaczynających się w zbiorze $\{x, z\}$ i kończących się w zbiorze $\{y, w\}$. Jeśli ścieżka wychodząca z x kończy się w y, to krawędzie grafu R^d leżące na lewo od tej ścieżki wyznaczają zbiór spójny łączący górny bok z dolnym. Wówczas zachodzi zdarzenie $V(R^d)$. Jeśli natomiast ścieżka wychodząca z x kończy się w w, to krawędzie grafu Rleżące na prawo od tej ścieżki wyznaczają zbiór spójny łączący lewy bok z prawym. Wówczas zachodzi zdarzenie H(R) (patrz Rysunek 17).

Musimy jeszcze udowodnić, że zdarzenia H(R) i $V(R^d)$ nie mogą zajść jednocześnie. Stwierdzenie to jest w zasadzie intuicyjnie bardzo jasne, jednak jego formalne uzasadnienie wymaga np. twierdzenia o krzywej Jordana. Jeśli np. istnieje ścieżka



Rysunek 16: Grafy R, R^d oraz graf pomocniczy (białe wierzchołki) ze zorientowanymi krawędziami.

z x do y, to idąc lewą krawędzią prostokąta \mathbb{R}^d możemy ją uzupełnić do krzywej Jordana γ (Rysunek 18). Przypuśćmy, że istnieje ścieżka łącząca lewą krawędź prostokąta z prawą. Wówczas Krzywa γ obiega punkt początkowy tej ścieżki dokładnie raz, a punkt końcowy zero razy. Zatem punkt końcowy znajduje się poza obszarem ograniczonym krzywą γ , czyli nasza ścieżka (z twierdzenia o krzywej Jordana!) musi przeciąć krzywą γ . Oznacza to, że pewna krawędź grafu \mathbb{R} przecina krawędź grafu pomocniczego, co przeczy konstrukcji tego grafu. Otrzymana sprzeczność kończy dowód.

Zauważmy, że jeśli krawędź kraty jest otwarta z prawdopodobieństwem p, to przecinająca ją krawędź kraty dualnej jest otwarta z prawdopodobieństwem 1 - p. Zatem z Lematu 8 wynika następujący wniosek.

Wniosek 1. Niech R będzie prostokątem k na l-1 i R' prostokątem k-1 na l.



Rysunek 17: Ścieżka wychodząca z punktu x i kończąca się w punkcie y oraz wyznaczony przez nią zbiór krawędzi w grafie dualnym R^d .

Wówczas

$$\mathbb{P}_p(H(R)) + \mathbb{P}_{1-p}(V(R')) = 1.$$

Jeśli k = l = n+1, to R i R' są obrucone względem siebie o 90°, a zatem $\mathbb{P}_p(H(R)) = \mathbb{P}_p(V(R'))$. Zatem w tym przypadku $\mathbb{P}_{1/2}(H(R)) = \mathbb{P}_{1/2}(V(R')) = 1/2$. Niech $S = [0, n] \times [0, n]$. Wtedy oczywiście $\mathbb{P}_p(H(R)) \leq \mathbb{P}_p(H(S)) = \mathbb{P}_p(V(S))$, a zatem

$$\mathbb{P}_{1/2}(H(S)) = \mathbb{P}_{1/2}(V(S)) \ge 1/2.$$

4.3 Dowód twierdzenia Harrisa

W tym rozdziale udowodnimy, że $p_c \geq 1/2$. Rozpoczniemy od następującego lematu.

Lemat 9. Niech $m \ge n$ i niech $S = [0, n] \times [0, n]$ oraz $R = [0, m] \times [0, 2n]$. Niech X(R) oznacza następujące zdarzenie: istnieją ścieżki P_1 i P_2 złożone z otwartych



Rysunek 18: Zamknięta krzywa Jordana wyznaczona przez ścieżkę od x do y.

krawędzi takie, że P_1 leży w S i łączy górny i dolny bok S oraz P_2 leży w R i łączy pewien punkt ścieżki P_1 z prawym bokiem R. Wówczas

$$\mathbb{P}_p(X(R)) \ge \frac{1}{2} \mathbb{P}_p(H(R)) \mathbb{P}_p(V(S)).$$

Dowód. Niech LV(S) będzie wertykalnym cięciem S położonym najbardzie po lewej stronie S. Niech P_1 będzie ustalonym cięciem wertykalnym S. Wówczas zdarzenie $\{LV(S) = P_1\}$ nie zależy od krawędzi położonych na prawo od ścieżki $P_1.$ Rozważmy ścieżkę \tilde{P}_1 będącą symetrycznym obrazem ścieżki P_1 w symetrii względem prostej $\{y = n\}.$ Oczywiście jeśli P_1 jest ścieżką otwartą, ścieżka \tilde{P}_1 nie musi być otwarta. Niech P będzie ścieżką powstałą przez połączenie ścieżek P_1 i \tilde{P}_1 (Rysunek 19). Każde horyzontalne cięcie R musi przecinać ścieżkę P (argument podobny do dowodu drugiej części Lemat 8). Ponieważ P jest symetryczna względem prostej $\{y = n\}$, zdarzenie polegające na istnieniu cięcia horyzontalnego R przecinającego P po raz ostatni w punkcie należącym do krzywej P_1 ma prawdopodobieństwo przynajmniej $\frac{1}{2}\mathbb{P}_p(H(R)).$ Niech

 $Y(P_1) = \{ \text{istnieje ścieżka } P_2 \le R \text{ na prawo od } P \text{ łącząca } P_1 \text{ z prawym bokiem } R \}.$

Wówczas z powyższego mamy $\mathbb{P}_p(Y(P_1)) \geq \frac{1}{2}\mathbb{P}_p(H(R))$. Zdarzenie $Y(P_1)$ zależy tylko od krawędzi leżących na prawo od P, a zatem zdarzenia $\{LV(S) = P_1\}$ i $Y(P_1)$ są niezależne. Stąd

$$\mathbb{P}_p\left(Y(P_1) \mid LV(S) = P_1\right) = \mathbb{P}_p\left(Y(P_1)\right) \ge \frac{1}{2}\mathbb{P}_p\left(H(R)\right),$$



Rysunek 19: Prostokąt Ri kwadrat S wraz ze ścieżkami P_1 , \tilde{P}_1 oraz P_2 .

gdzie P_1 jest ustalone. Zauważmy, że zdarzenia $\{LV(S)=P_1\}$ i $Y(P_1)$ implikują X(R). Zatem

$$\frac{1}{2}\mathbb{P}_p\left(H(R)\right) \le \mathbb{P}_p\left(Y(P_1) \mid LV(S) = P_1\right) = \frac{\mathbb{P}_p\left(Y(P_1) \text{ i } LV(S) = P_1\right)}{\mathbb{P}_p\left(LV(S) = P_1\right)} \le \frac{\mathbb{P}_p\left(X(R) \text{ i } LV(S) = P_1\right)}{\mathbb{P}_p\left(LV(S) = P_1\right)}.$$

Stąd, uwzględniając fakt, że $V(S) = \bigcup_{P_1} \{ LV(S) = P_1 \}$ i jest to suma rozłączna,

$$\frac{\mathbb{P}_p\left(X(R)\right)}{\mathbb{P}_p\left(V(S)\right)} \ge \frac{\mathbb{P}_p\left(X(R) \text{ i } V(S)\right)}{\mathbb{P}_p\left(V(S)\right)} = \sum_{P_1} \frac{\mathbb{P}_p\left(X(R) \text{ i } LV(S) = P_1\right)}{\mathbb{P}_p\left(V(S)\right)}$$
$$\ge \frac{1}{2} \mathbb{P}_p\left(H(R)\right) \sum_{P_1} \frac{\mathbb{P}_p\left(LV(S) = P_1\right)}{\mathbb{P}_p\left(V(S)\right)} = \frac{1}{2} \mathbb{P}_p\left(H(R)\right).$$

Korzystając z Lematu 9 udowodnimy następujący lemat.

Lemat 10. Niech $\rho > 1$ bedzie liczbą naturalną. Istnieje stała $c_2(\rho) > 0$ taka, że dla $R = [0, 2\rho n] \times [0, 2n]$ mamy

$$\mathbb{P}_{1/2}(H(R)) \ge c_2(\rho).$$

Dowód. Dla $R = [0,m] \times [0,n]$ niech $h_{m,n} = \mathbb{P}_{1/2}(H(R))$. Udowodnimy nierówność

$$h_{2m-n,2n} \ge \frac{1}{2^5} h_{m,2n}^2, \qquad m \ge n.$$
 (9)

Z nierówności (9) wynika już teza lematu. Faktycznie, ponieważ $h_{2n,2n} \ge 1/2$ na mocy Wniosku 1, bez trudu udowadniamy przez indukcję nierówność $h_{(2^l+1)n,2n} \ge \frac{1}{2^{6\cdot 2^l-5}}$ dla $l \ge 0$. Niech $l_0(\rho) = \inf\{l: 2^l + 1 \ge 2\rho\}$. Wówczas

$$h_{2n\rho,2n} \ge h_{(2^{l_0}+1)n,2n} \ge \frac{1}{2^{6 \cdot 2^{l_0}-5}} \ge \frac{1}{2^{24\rho-5}}$$

gdyż $2^{n_0-1} < 2^{n_0-1} + 1 < 2\rho$.



Rysunek 20: Prostokąty R i R' oraz zdarzenia E_1 i E_2 .

Przystępujemy do dowodu nierówności (9). Niech $R = [0, m] \times [0, 2n]$ i niech $R' = [n - m, n] \times [0, 2n]$. Niech $E_1 = X(R)$ i niech E_2 będzie odbitym zdarzeniem X(R') (patrz Rysunek 20). Dodatkowo dla $S = [0, n] \times [0, n]$ niech $E_3 = H(S)$. Oczywiście zdarzenia E_1, E_2, E_3 są zdarzeniami monotonicznymi oraz ich koniunkcja implikuje zdarzenie $H(R \cup R')$. Stąd z Lematu 7 mamy

$$\mathbb{P}_{1/2}(H(R \cup R')) \ge \mathbb{P}_{1/2}(E_1 \cap E_2 \cap E_3) \ge \mathbb{P}_{1/2}(E_1) \mathbb{P}_{1/2}(E_2) \mathbb{P}_{1/2}(E_3)$$

= $\mathbb{P}_{1/2}(X(R))^2 \mathbb{P}_{1/2}(H(S)).$

z Lematu 9 mamy

$$\mathbb{P}_{1/2}(X(R)) \ge \frac{1}{2} \mathbb{P}_{1/2}(H(R)) \mathbb{P}_{1/2}(V(S))$$

Ponadto z Wniosku 1 mamy $\mathbb{P}_{1/2}(V(S)) = \mathbb{P}_{1/2}(H(S)) \ge 1/2$. Stąd

$$\mathbb{P}_{1/2}\left(H(R \cup R')\right) \ge \frac{1}{4} \mathbb{P}_{1/2}\left(H(R)\right)^2 \mathbb{P}_{1/2}\left(V(S)\right)^2 \mathbb{P}_{1/2}\left(H(S)\right) \ge \frac{1}{2^5} \mathbb{P}_{1/2}\left(H(R)\right)^2.$$



Rysunek 21: Prostokąty R_1 , R_2 , R_3 i R_4 oraz odpowiadające im cięcia $H(R_1), V(R_2), H(R_3), V(R_4)$.

Oczywiście $R \cup R'$ ma szerokość 2m - n i wysokość 2n, a zatem udowodniliśmy nierówność (9).

Dowód twierdzenia Harrisa. Rozważmy pierścień ograniczony dwoma kwadratami o bokach długości 6*n* oraz 2*n*. Pierścień ten da się pokryć czterema prostokątami R_1, R_2, R_3, R_4 . Zauważmy, że zdarzenia $H(R_1), V(R_2), H(R_3)$ i $V(R_4)$ implikują istnienie otwartego cyklu otaczającego środek pierścienia (Rysunek 21). Z Lematu 7 i z Lematu 10 dla $\rho = 3$ mamy

$$\mathbb{P}_{1/2}(H(R_1) \cap V(R_2) \cap H(R_3) \cap V(R_4)) \ge \mathbb{P}_{1/2}(H(R_1))^4 \ge c_2^4,$$

gdzie $c_2 = c_2(3)$. Rozważmy ciąg rozłącznych pierścieni A_1, A_2, \ldots o środkach w 0, gdzie A_k jest ograniczony kwadratami o bokach 4^k i $3 \cdot 4^k$. Niech C_k oznacza zdarzenie polegające na tym, że pierścień A_k nie zawiera cyklu otaczającego środek tego pierścienia. Zdarzenie to ma prawdopodobieństwo nie większe niż $1 - c_2^4 < 1$. Zdarzenia C_1, C_2, \ldots są niezależne, a zatem ich koniunkcja ma prawdopodobieństwo 0. Wynika stąd, że z prawdopodobieństwem 1 istnieje cykl \mathcal{C} otaczający punkt (1/2, 1/2) na kracie dualnej. Wynika stąd, że spójna składowa punktu (1/2, 1/2) w kracie dualnej jest skończona i ograniczona cyklem \mathcal{C} . Stąd $\theta(1/2) = 0$, a zatem $\mathbb{P}_{1/2}(E_{\infty}) = 0$. \Box

4.4 Dowód twierdzenia Kestena

Rozpoczniemy od następującego lematu, mówiącego, że dla p > 1/2 prawdopodobieństwo istnienia otwartego cięcia dla dużych prostokątów $12n \times 4n$ jest bardzo bliskie 1.



Rysunek 22: Dyskretny torus \mathbb{T}_{20} wraz z zaznaczonym prostokątem 5×3 .

Lemat 11. Dla p > 1/2 istnieją stał
e $\delta = \delta(p) > 0$ i $n_0 = n_0(p)$ takie, że dla
 $n \ge n_0$ i prostokąta R_n o szerokości
12n i wysokości 4n mamy

$$\mathbb{P}_p(H(R_n)) \ge 1 - n^{-\delta}.$$

Dowód. Jeśli R jest prostokątem 14n na 2n to $\mathbb{P}_{1/2}(H(R)) \ge c_2$ dla stałej $c_2 = c_2(7)$ z Lematu 8. Aby jednak móc skorzystać z Twierdzenia 1, musimy skonstruować odpowiednie symetryczne zdarzenia losowe. W tym celu rozważmy dyskretny torus \mathbb{T}_n , czyli zbiór $\mathbb{Z}^2/(n\mathbb{Z} \times n\mathbb{Z})$, przy czym wierzchołki (k_1, l_1) i (k_2, l_2) łączymy krawędzią jeśli $|k_1 - k_1| \equiv 1 \mod n$ oraz $l_1 - l_2 \equiv 0 \mod n$ lub $|l_1 - l_1| \equiv 1 \mod n$ oraz $k_1 - k_2 \equiv 0 \mod n$. Mamy zatem n^2 wierzchołków i $N = 2n^2$ krawędzi (Rysunek 22).

Niech $1 \leq k, l \leq n-2$. Prostokątem $k \times l \le \mathbb{T}_n$ nazywamy graf indukowany z podgrafu \mathbb{Z}^2 będącego prostokątem $k \times l \le t$ w tym grafie. Ponieważ $k, l \leq n-2$, prostokąty k na $l \le \mathbb{T}_n$ są izomorficzne z prostokątami $k \times l \le \mathbb{Z}^2$. Są one zbyt małe, żeby mogły się na torusie zawinąć. Innymi słowy, mamy do czynienia z normalnymi prostokątami, ale narysowanymi na torusie. Każda krawędź se \mathbb{T}_n jest otwarta z prawdopodobieństwem p i zamknięta z prawdopodobieństwem 1-p, oczywiście niezależnie dla każdej krawędzi. Odpowiadającą temu modelowi miarę na przestrzeni konfiguracji krawędzi $\{O, Z\}^N$ oznaczamy przez $\mathbb{P}_p^{\mathbb{T}_n}$. Niech E_n bedzie zdarzeniem polegającym na tym, że na \mathbb{T}_n istnieje prostokąt $14n \times 2n$ z otwartym cięciem horyzontalnym lub prostokąt $2n \times 14n$ z otwartym cięciem wertykalnym. Zauważmy, że E_n jest zdarzeniem symetrycznym. Dla $\mathbb{T} = \mathbb{T}_{16n}$ liczba krawędzi wynosi $N = 512n^2$. Niech R będzie ustalonym prostokątem $14n \times 2n$ na \mathbb{T} i niech R' bedzie odpowiadającym mu prostokątem na \mathbb{Z}^2 . Wówczas

$$\mathbb{P}_{1/2}^{\mathbb{T}}(E_n) \ge \mathbb{P}_{1/2}^{\mathbb{T}}(H(R)) = \mathbb{P}_{1/2}(H(R')) \ge c_2.$$

Niech $\delta = \frac{p-\frac{1}{2}}{64c_1}$, gdzie c_1 jest stałą z Twierdzenia 1 i niech $\varepsilon = n^{-128\delta}$. Istnieje $n_0 = n_0(p)$ takie, że dla $n \ge n_0$ mamy $\varepsilon < c_2 \le 1/2$. Oczywiście

$$p - \frac{1}{2} = 64c_1\delta = c_1 \frac{\ln(1/\varepsilon)}{\ln(n^2)} > c_1 \frac{\ln(1/2\varepsilon)}{\ln(512n^2)}.$$

Korzystając z notacji użytej w Twierdzeniu 1 wprowadzamy liczby $p_{-}(\varepsilon), p_{+}(\varepsilon) \in [0, 1]$ spełniające równości

$$\mathbb{P}_{p_{-}(\varepsilon)}^{\mathbb{T}}(E_{n}) = \varepsilon, \qquad \mathbb{P}_{p_{+}(\varepsilon)}^{\mathbb{T}}(E_{n}) = 1 - \varepsilon.$$

Ponieważ $\mathbb{P}_{1/2}^{\mathbb{T}}(E_n) \geq c_2 > \varepsilon$, mamy $p_-(\varepsilon) < 1/2$. Korzystając z Twierdzenia 1 otrzymujemy

$$p_{+}(\varepsilon) - \frac{1}{2} < p_{+}(\varepsilon) - p_{-}(\varepsilon) < c_{1} \frac{\ln(1/2\varepsilon)}{\ln(512n^{2})} < p - \frac{1}{2}$$

a zatem $p > p_+(\varepsilon)$. Wynika stąd, że

$$\mathbb{P}_p^{\mathbb{T}}(E_n) \ge 1 - \varepsilon = 1 - n^{-128\delta}.$$



Rysunek 23: Prostokąty R i R_i oraz ich wzajemne położenie w zależności od współrzędnej lewego dolnego rogu prostokąta R (czarna kropka). Litera n oznacza współrzedną nieparzystą zaś litera p współrzędną parzystą.

Musimy teraz przejść od zdarzenia E_n do zdarzenia H(R). Rozważmy pokrycie torusa $\mathbb{T} = \mathbb{T}_{16n}$ za pomocą 64 prostokątów R_1, \ldots, R_{64} o wymiarach $12n \times 4n$, z których każdy ma lewy dolny róg w punkcie (k, l) o współrzędnych parzystych. Oczywiście niektóre prostokąty przecinają się wzajemnie. Zauważmy, że dla dowolnego prostokąta R o wymiarach $14n \times 2n$ istnieje $1 \leq i \leq 64$ takie, że $R \cap R_i$ jest prostokątem o wymiarach $12n \times 2n$ (Rysunek 23). W tym przypadku H(R) implikuje $H(R_i)$. Pododnie, istnieją prostokąty R_{65}, \ldots, R_{128} wymiarów $4n \times 12n$ pokrywające \mathbb{T} takie, że dla każdego prostokąta R wymiarów $2n \times 14n$ istnieje $65 \leq i \leq 128$ takie, że $R \cap R_i$ jest prostokątem o wymiarach $2n \times 12n$. W tym przypadku V(R) implikuje $V(R_i)$. Niech $E_{n,i} = H(R_i)$ dla $1 \leq i \leq 64$ oraz $E_{n,i} = V(R_i)$ dla $65 \leq i \leq 128$. Z powyższego wynika, że jeśli E_n zachodzi to zachodzi również jedno ze zdarzeń $E_{n,i}$, $1 \leq i \leq 128$. Przechodząc do dopełnień mamy

$$\bigcap_{i=1}^{128} E_{n,i}^c \subseteq E_n^c.$$

Zdarzenia $E_{n,i}^c$ są zdarzeniami malejącymi. Stąd na mocy Lematu 7 mamy

$$\mathbb{P}_p^{\mathbb{T}}(E_n^c) \ge \mathbb{P}_p^{\mathbb{T}}\left(\bigcap_{i=1}^{128} E_{n,i}^c\right) \ge \prod_{i=1}^{128} \mathbb{P}_p^{\mathbb{T}}(E_{n,i}^c) = \mathbb{P}_p^{\mathbb{T}}(E_{n,1}^c)^{128}.$$

Zatem

$$\mathbb{P}_{p}^{\mathbb{T}}(E_{n,1}^{c}) \leq \mathbb{P}_{p}^{\mathbb{T}}(E_{n}^{c})^{\frac{1}{128}} = \left(1 - \mathbb{P}_{p}^{\mathbb{T}}(E_{n})\right)^{\frac{1}{128}} \leq n^{-\delta}$$

Stąd

$$\mathbb{P}_p^{\mathbb{T}}(E_{n,1}) \ge 1 - n^{-\delta}.$$

Zdarzenie $E_{n,1}$, polegające na istnieniu horyzontalnego cięcia dla ustalonego prostokąta $12n \times 2n$ na torusie, może być utożsamione ze zdarzeniem polegającym na istnieniu analogicznego cięcia dla ustalonego prostokąta $12n \times 2n$ na kracie \mathbb{Z}^2 . Zatem

$$\mathbb{P}_p(H(R)) = \mathbb{P}_p^{\mathbb{T}}(E_{n,1}) \ge 1 - n^{-\delta}.$$

Oczywiście z powyższego lematu wynika analogiczny lemat dla prostokątów $8n \times 4n$.

Lemat 12. Dla p > 1/2 istnieją stał
e $\delta = \delta(p) > 0$ i $n_0 = n_0(p)$ takie, że dla
 $n \ge n_0$ i prostokąta R_n o szerokości
 8ni wysokości 4n mamy

$$\mathbb{P}_p(H(R_n)) \ge 1 - n^{-\delta}.$$

Jesteśmy gotowi to udowodnienia twierdzenia Kestena.

Dowód twierdzenia Kestena. Niech p > 1/2. Niech $\delta = \delta(p)$ i $n_0 = n_0(p)$ będą dane przez Lemat 12. Niech $m \ge n_0$ i niech n = 4m. Dla $k = 0, 1, \ldots$ definiujemy prostokąty $R_k = [0, 2^k n] \times [0, 2^{k+1} n]$ gdy k jest parzyste i $R_k = [0, 2^{k+1} n] \times [0, 2^k n]$ gdy k jest nieparzyste (Rysunek 24).

Niech $E_k = H(R_k)$ dla k nieparzystych i $E_k = V(R_k)$ dla k parzystych. Jeśli zachodzą wszystkie zdarzenia $E_k, k \ge 0$ to zachodzi również zdarzenie E_{∞} (cięcia horyzontalne i wertykalne dla prostokątów R_k i R_{k+1} muszą się spotkać). Stąd

$$\mathbb{P}_p\left(E_{\infty}^c\right) \le \mathbb{P}_p\left(\bigcup_{k\ge 0} E_k^c\right) \le \sum_{k\ge 0} \mathbb{P}_p\left(E_k^c\right) \le \sum_{k\ge 0} (2^k n)^{-\delta} = \frac{n^{-\delta}}{1-2^{-\delta}} < 1,$$

o ile n = 4m jest dostatecznie duże. Wynika stąd, że $\mathbb{P}_p(E_{\infty}) > 0$.



Rysunek 24: Prostokąty R_0 , R_1 , R_2 i R_3 oraz odpowiadające im otwarte cięcia E_0 , E_1 , E_2 oraz E_3 .
Literatura

- BR B. Bollobás, O. Riordan, A short proof of the Harris-Kesten theorem.
- [ER] P. Erdös, and A. Rényi, On random graphs I, Publ. Math. Debrecen 6 (1959), 290–297.
- [I] E. Ising, Zeitschrift f. Physik 31, 253 (1925).
- [O] L. Onsager, Phys. Rev. 65, 117 (1944).
- [JS] J. Jakubowski, R. Sztencel, Wstęp do teorii prawdopodobieństwa, wyd. III, Script, Warszawa 2004.
- [F] R. P. Feynman, Statistical Mechanics. A set of lectures. The Benjamin and Cummings Publishing Co., 1972.
- [KF] E. Friedgut, G. Kalai, Every monotone graph property has a sharp threshold, Proc. Amer. Math. Soc. 124 (1996) 2993–3002.
- [G] G. Grimmett, Percolation, 2nd edn (Springer, Berlin, 1999).
- [H] T. E. Harris, A lower bound for the critical probability in a certain percolation process', Proc. Cambr. Philos. Soc. 56 (1960) 13–20.
- [K] H. Kesten, The critical probability of bond percolation on the square lattice equals 1/2', Comm. Math. Phys. 74 (1980) 41–59.
- [KW] M. Kac and J. C. Ward, Phys. Rev. 88, 1332 (1952).
- [LL] Landau, L.D., Lifshitz, E.M. (1972). Mechanics Course of Theoretical Physics, Vol. 1. Franklin Book Company. ISBN 0-08-016739-X.
- [Be] W. Beckner, *Inequalities in Fourier analysis*, Annals of Math. 102 (1975), 159–182.
- [Bo] A. Bonami, Etude des coefficients Fourier des fonctiones de $L_p(G)$, Ann. Inst. Fourier 20 (1970), 335–402.
- [G1] L. Gross, Logarithmic Sobolev inequalities, Amer. J. Math. 97 (1975) 1061–1083.
- [KKL] J. Kahn, G. Kalai and N. Linial, The influence of variables on Boolean functions, in Proc. 29-th Annual Symposium on Foundations of Computer Science, 68-80, 1988.

- [BKKKL] J. Bourgain, J. Kahn, G. Kalai, Y. Katznelson and N. Linial, The influence of variables in product spaces, Israel J. Math. 77 (1992) 55–64.
- [T] M. Talagrand, On Russo's approximate zero-one law. Ann. Probab., 22(3):1576-1587, 1994.

Piotr Nayar Institute of Mathematics, University of Warsaw, Banacha 2, 02-097 Warszawa, Poland. nayar@mimuw.edu.pl