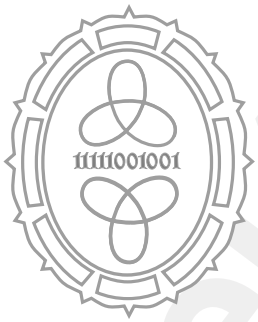
UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU
WYDZIAŁ MATEMATYKI I INFORMATYKI



Filip Piękniewski
nr albumu: 146961

Analiza stabilności i diagramów fazowych dla lokalnie hopfieldowskich sieci neuronowych w obecności wysoce skorelowanych wzorców okresowych

Praca magisterska na kierunku:
INFORMATYKA
Promotor:
dr Tomasz Schreiber,
Katedra Teorii Prawdopodobieństwa i Analizy Stochastycznej

Toruń 2005

Tylko do celów edukacyjnych

Słowa kluczowe

Miary Gibbsa, rekurencyjne sieci neuronowe, Łańcuchy Markowa, teoria Pirogova-Sinaia, algorytmy stochastyczne, minimalizacja globalna, rozkłady stacjonarne, diagramy fazowe, przejście fazowe, model Isinga

Tylko do celów edukacyjnych

Spis treści

1. Wstęp 4
1.1. Sformułowanie problemów informatycznych w języku minimalizacji globalnej 5
1.2. Analogie informatyki i fizyki - przestrzeń problemu - przestrzeń fazowa 6
1.3. Problem minimów lokalnych - wyżarzanie 7
1.4. Inspiracje biologiczne w informatyce - sieci neuronowe, perceptrony, MLP . . 8
1.5. Przegląd algorytmów minimalizacji globalnej 10
1.5.1. Algorytm Random Walk 10
1.5.2. Algorytm Tabu-Search 10
1.5.3. Algorytm wielu startów (multistart) 11
1.5.4. Algorytm P* 11
1.5.5. Simulated Annealing 11

2. Sieci rekurencyjne, sieci Hopfielda 12
2.1. Definicja, przykłady dynamiki 12
2.2. Sieci Hopfielda, pojemność pamięciowa, katastrofa zapominania 14
2.3. Przykłady problemów przetłumaczonych na język sieci rekurencyjnych 18
2.3.1. Problem kolorowania grafu 18
2.3.2. Problem planu zajęć 19
2.3.3. Problem komiwojażera 20
2.4. Elementy maszyn Boltzmanna 21

3. Łańcuchy Markowa 25
3.1. Definicja i własności łańcuchów Markowa 25
3.2. Rozkłady stacjonarne 27
3.3. Algorytm doskonały Proppa-Wilsona 32

4. Aparat termodynamiki 35
4.1. Przykłady układów opisywanych za pomocą fizyki statystycznej 36
4.1.1. Klasyczny gaz 36
4.1.2. Gaz konfiguracyjny - model matematyczny 37
4.1.3. Zbiór spinów magnetycznych, model Isinga 38
4.2. Rozkłady mikrokanoniczne i kanoniczne (miary Gibbsa) 38
4.3. Miary graniczne Gibbsa w sensie klasycznym i DLR 41
4.4. Struktura zbioru miar DLR, pojęcie przejścia fazowego 43
4.5. Argument Peierlsa - przejście fazowe w modelu Isinga 44
4.6. Elementy teorii Pirogova-Sinaia 47

5. Model lokalnie Hopfieldowskiej sieci neuronowej 50
5.1. Pojęcia wstępne 50
5.2. Opis modelu 50
5.3. Wyniki obliczeń i ich związek z przewidywaniami teoretycznymi 54
5.4. Potencjalne możliwości prowadzenia dalszych badań 61

TylkoDoCelówEdukacyjnych

6. Dodatki 62
6.1. Szczegóły implementacyjne programu 62
6.2. Płytka CD z filmami zawierającymi ewolucję układu 63
6.3. Podziękowania 63

Spis tabel 68

Spis rysunków 69

Skorowidz 70

Tylko do celów edukacyjnych

ROZDZIAŁ 1

Wstęp

Jednym z fascynujących zagadnień dzisiejszej informatyki jest analiza różnorodnych modeli obliczeń. W tym kontekście na szczególną uwagę zasługują modele zakładające równoległość, lub wręcz masywną równoległość obliczeń. Zagadnienia te, chociaż nie są wcale nowe, wydają się mieć teraz kluczowe znaczenie, ze względu na kończące się możliwości ciągłego przyspieszania urządzeń działających w modelu sekwencyjnym (wiele wskazuje na to, że jesteśmy blisko granic możliwości zwiększania częstotliwości procesorów, możemy jednak nadal zwiększać liczbę procesorów). Okazuje się, że pojęcie obliczenia nie musi być zarezerwowane dla układów arytmetycznych, także układy o zupełnie innych własnościach niż typowe bramki logiczne, mogą wykonywać efektywnie skomplikowane zadania (czego najlepszym przykładem jest chociażby mózg). Badania w dziedzinie klasycznych algorytmów równoległych mają swój udział w tym, że dzisiaj umiemy za pomocą komputerów równoległych całkiem trafnie przewidywać numerycznie pogodę, czy własności konstrukcyjne budynków, samolotów itp. Dzięki badaniom nad układami masywnie równoległymi, takimi jakimi są na przykład sieci neuronowe mamy dzisiaj wiele doskonałych czujników, układów przetwarzania obrazu a także znacznie większe pojęcie na temat tego jak działa ludzki mózg. Nie da się jednak powiedzieć, że wiemy już jak dokładnie działają skomplikowane układy składające się z tysięcy prostych elementów takich jak neurony. W tym miejscu warto nadmienić o nieuchronnym zbliżeniu informatyki z fizyką. Przejawia się to zarówno w rosnącym znaczeniu pojęcia informacji w zrozumieniu praw rządzących Wszechświatem, jak i zainteresowanie zagadnieniem komputerów kwantowych, kryptografii kwantowej itp. . Istotny udział ma także wykorzystywanie aparatu matematycznego wywodzącego się z fizyki do opisu pewnych teorii informatycznych. Przykładowo dziedziną która zajmuje się opisem skomplikowanych układów w których wiele rzeczy dzieje się równolegle to fizyka statystyczna. Jej aparat matematyczny był dotychczas stosowany do opisu układów fizycznych takich jak gaz doskonały czy zbiór spinów magnetycznych. W kontekście sieci neuronowych stosowanie aparatu fizyki statystycznej jest obecne między innymi w badaniu zjawisk występujących dla wielkoskalowych sieci rekurencyjnych (szkła spinowe, model Hopfielda itp.). W tej pracy prezentowane jest nieco inne podejście do zastosowania aparatu fizyki statystycznej dla sieci neuronowych. Okazuje się, że w pewnych układach (sieciach Hopfielda o zlokalizowanej strukturze połączeń) dobry opis teoretyczny zapewnia teoria Pirogova-Sinaia opisująca własności diagramów fazowych dla kolektywnych układów z oddziaływaniami zadanymi przez potencjały lokalne. Teoria ta rozwinięta w latach 80 XX wieku jest jeszcze dość nowa, bardzo skomplikowana, być może jednak w jakieś mierze stanowi wskazówkę do zrozumienia własności skomplikowanych układów neuronów. Praca ta zawiera wprowadzenie do wszystkich zagadnień potrzebnych do zrozumienia wyników teorii Pirogova-Sinaia (dowody opisywanych wyników wykraczają jednak daleko poza zakres bieżącej rozprawy), oraz rezultat eksperymentu obliczeniowego, który potwierdza słuszność zastosowania tej teorii do opisu lokalnie Hopfieldowskich sieci neuronowych. Przedstawione tu wyniki zostały przyjęte do prezentacji na konferencji International Joint Conference on Neural Networks 2005 odbywającej się w Montrealu [30].

Tylko do celów edukacyjnych

1.1. Sformułowanie problemów informatycznych w języku minimalizacji globalnej

Komputery spełniają dzisiaj wiele zadań, począwszy od pomagania w projektowaniu, poprzez obliczenia naukowe, skończywszy na rozrywce. W rzeczywistości każde z tych zagadnień rozbijane jest na tysiące prostych czynności które komputer wykonuje niezwykle szybko. Na tak elementarnym poziomie wykonywane są właściwie jedynie operacje arytmetyczne i zapisywanie/odczytywanie pamięci a wiele jednakowych operacji można wykonać w na różne sposoby. Niski poziom analizy działania komputera jest ciekawy z punktu widzenia teorii obliczeń (konstrukcja maszyn Turinga, lingwistyka itp), jednak w tym tekście będziemy operować na bardziej ogólnych pojęciach. Do dalszych rozważań musimy zdefiniować pewne pomocnicze pojęcie, które pozwoli opisać czym z punktu widzenia rozważań tu prezentowanych będzie typowe zagadnienie rozwiązywane przez komputer.

Zagadnieniem obliczeniowym nazwiemy każdą trójkę $(S, \mathbb{D}, \mathfrak{R})$, gdzie S jest specyfikacją problemu (sformułowaną w pewnym języku formalnym, lub poprzez zadanie maszyny obliczeniowej), $\mathbb{D}$ zbiorem możliwych danych wejściowych a $\mathfrak{R}$ zbiorem wszystkich możliwych rozwiązań problemu. S specyfikuje funkcję która dla pewnych danych wejściowych $D \in \mathbb{D}$ przyjmuje wartość będącą podzbiorem $R \subset \mathfrak{R}$ (podzbiór zbioru możliwych rozwiązań[1]). Tak ogólne pojęcie mieści w sobie oczywiście klasyczne problemy, takie jak na przykład znalezienie cyklu Hamiltona w grafie, czy wykonanie BFS. W pierwszym przypadku rozwiązaniem jest zbiór wszystkich cykli Hamiltona dla danego grafu, w drugim drzewo rozpięte przez przeszukiwanie BFS. Do przedstawionego schematu pasują również mniej klasyczne problemy takie jak np. rozpoznanie litery na zaszumionej bitmapie, czy rozpoznanie mowy. Niezależnie od charakteru problemu, można wyróżnić pewne cechy wspólne:

- Dla każdych danych wejściowych $D \in \mathbb{D}$ istnieje podzbiór R zbioru $\mathfrak{R}$ stanowiący rozwiązanie zagadnienia obliczeniowego (który może być podzbiorem pustym - jeśli dla danego wejścia problem nie posiada rozwiązania[2])

- Dla każdego zagadnienia obliczeniowego istnieje odwzorowanie między zbiorem danych $\mathbb{D}$ a zbiorem podzbiorów zbioru dopuszczalnych rozwiązań $2^{\mathfrak{R}}$[3]. Każdym danym wejściowym odpowiada pewien podzbiór zbioru rozwiązań (jedno elementowy, gdy rozwiązanie jest jednoznaczne)

Z powyższych punktów wynika, że można przeformułować zadanie rozwiązania zagadnienia obliczeniowego w następujący sposób:

- Definiujemy pewną funkcję $\phi_D : \mathfrak{R} \to \mathbb{R}$ ograniczoną z dołu taką, że dla pewnych danych D:

$$argmin_{r \in \mathfrak{R}} \phi_D(r) = \{r \in \mathfrak{R}; \text{r jest rozwiązaniem S przy danych D}\} \tag{1.1}$$

  Zatem funkcja ta osiąga swoje minima dla elementów $r \in R$ które należą do rozwiązania problemu S dla danych D. Trzeba tu dodać, że w przypadku, gdy dane zagadnienie nie posiada rozwiązania ($R = \emptyset$), funkcja ϕ_D może się minimalizować na pewnych

[1] Element zbioru $\mathfrak{R}$ będziemy nazywać *rozwiązaniem możliwym*, w przeciwieństwie do *rozwiązania zagadnienia* (konkretnego).

[2] Aby uniknąć nieporozumień ustalmy, że *rozwiązaniem zagadnienia* nazywamy właśnie wspomniany podzbiór zbioru *możliwych rozwiązań* $\mathfrak{R}$, nie zaś element zbioru $\mathfrak{R}$.

[3] Przez $\mathfrak{R}$ oznaczamy zbiór wszystkich możliwych rozwiązań problemu

innych elementach zbioru $\mathfrak{R}$. O ile takie rozwiązania mają sens[4], będziemy je nazywać *suboptymalnymi*.

- Zagadnienie rozwiązania problemu S dla danych D sprowadza się wtedy do znalezienia minimum (lub minimów) funkcji ϕ_D

Zauważmy przy tym, że funkcja ϕ może być dobrana z pewną dowolnością. Wymagamy od niej jedynie tego aby osiągała minimum w rozwiązaniach problemu. W szczególności może to być na przykład funkcja która dla danego grafu G i danego uporządkowania jego wierzchołków osiąga 0, gdy uporządkowanie to tworzy cykl Hamiltona i 1 w przeciwnym wypadku. W ogólności jeśli posiadamy efektywną procedurę weryfikacyjną dla problemu to od razu możemy z niej skonstruować odpowiednią funkcję analogicznie do powyższego schematu. Jeśli problem jest zagadnieniem aproksymacyjnym to za minimalizowaną funkcję wystarczy wziąć normę różnicy aproksymatora od funkcji aproksymowanej itp.

Pewną ciekawostką jest to, że komputery przez większość swojego czasu obliczeniowego nie rozwiązują żadnych zagadnień obliczeniowych nawet w tak ogólnym sensie. Ich działanie zdominowane jest przez *jałowe czynności* związane z zerowaniem lub przenoszeniem fragmentów pamięci, operacjami wejścia/wyjścia itp. Można jednak spojrzeć na to zagadnienie szerzej, jeśli bowiem np. rysowanie ładnych okienek nie wydaje się być rzeczywistym problemem informatycznym to może być ono interpretowane jako czynność mająca na celu zoptymalizowanie komunikacji dwóch systemów informatycznych (komputera i mózgu użytkownika) i w tym sensie też jest pewnym zadaniem optymalizacyjnym[5].

1.2. Analogie informatyki i fizyki - przestrzeń problemu - przestrzeń fazowa

Nawiązując do rozważań ze wstępu można wskazać szereg analogii pomiędzy pojęciami pochodzenia informatycznego a pojęciami pochodzenia fizycznego. W fizyce podstawą wszelkich ścisłych przemyśleń jest model pewnego zjawiska, opisany równaniami matematycznymi. Zjawisko fizyczne przebiega w czasie i w drodze swojej ewolucji przyjmuje pewne stany. Funkcję która dla pewnego stanu zjawiska determinuje jego następny stan nazywamy dynamiką, a układ wraz z taką funkcją nazywamy układem dynamicznym[6]. Pomocnym okazuje się spojrzenie na układ z nieco innej perspektywy, mianowicie nie na pojedynczy jego stan, ale na przestrzeń wszystkich możliwych stanów zwaną dalej przestrzenią fazową (przestrzenią faz - stanów). Dynamika zadaje oczywiście odwzorowanie przestrzeni fazowej w siebie. Każda ewolucja układu zgodna z jego dynamiką wyznacza trajektorie w przestrzeni fazowej, przy czym można wyróżnić cały szereg możliwych trajektorii, począwszy od takich które zbiegają do pewnego punktu (atraktora) i zatrzymują się (atraktor jest punktem niezmienniczym dynamiki), przez takie które wpadają w cykliczną pętle lub zestaw pętli (zachowanie okresowe i pseudo-okresowe), aż po takie które w nie regularny sposób wypełniają całą przestrzeń (zachowanie chaotyczne). Istotą analogii modeli fizycznych i informatycznych jest jednak co innego, mianowicie fakt, iż dla tych pierwszych definiuje się (zazwyczaj w dość naturalny sposób) pewną dodatkową funkcję - energię. W ewolucji układów fizycznych posiadających

[4] To może być niejasne, ale zależy to od rozważanego problemu - kiedy pewne rozwiązanie jeszcze uznamy za sensowne (choć nieoptymalne) a kiedy uznamy je za niepoprawne.

[5] Chociaż rozwiązywanym tak naprawdę raz, przez projektantów interejsu systemu.

[6] W tym sformułowaniu jest oczywiście pewna przesada, układ można nazywać układem dynamicznym, gdy spełnia pewne dodatkowe założenia, przywoływanie jednak tych szczegółów na tym poziomie abstrakcji nie wydaje się konieczne.

możliwość odprowadzenia nadmiaru energii (np. w postaci ciepła, promieniowania etc.) stabilne są stany równowagowe odpowiadające zwykle lokalnym minimom funkcji energetycznej. Układy te zatem same z siebie w sprzyjających warunkach[7] wykonują zadanie minimalizacji pewnej funkcji energii na przestrzeni fazowej. Jeśli teraz przyjmiemy, że przestrzeń fazowa odpowiada przestrzeni problemu i zadamy funkcję energetyczną w taki sposób aby przyjmowała minima globalne w punktach będących rozwiązaniem problemu, to zadanie informatyczne polegające na znalezieniu algorytmu rozwiązującego dane zagadnienie tłumaczy się na problem znalezienia odpowiedniego mechanizmu (minimalizującego zadaną energię) i pozwoleniu aby układ ewoluował. Wtedy (pomijając pewne szczegóły o których za chwile) po odpowiednio długim czasie ewolucja układu powinna skończyć się w interesującym nas rozwiązaniu.

1.3. Problem minimów lokalnych - wyżarzanie

Niestety omówione wyżej układy fizyczne nie potrafią w ogólności rozwiązywać zagadnienia minimalizacji globalnej, są bowiem podatne na możliwość trafienia w minimum lokalne z którego już nie będą w stanie wyjść. Jeśli układ jest deterministyczny, wtedy każde minimum lokalne funkcji energetycznej na przestrzeni fazowej jest atraktorem dynamiki. Tym samym każda trajektoria zaczyna się w pewnym basenie atrakcji pewnego minimum lokalnego, do którego prędzej lub później wpadnie (i już w nim zostanie). Może się oczywiście zdarzyć, że przypadkowo wybierzemy taki stan początkowy układu, który leży w basenie atrakcji minimum globalnego, jednak taka sytuacja jest w ogólności mało prawdopodobna[8]. Samo skonstruowanie układu rozwiązującego problem inspirowanego pojęciami rodem z fizyki nie wystarcza do jego rozwiązania. Potrzebne są dodatkowe warunki zadane na dynamikę, coś co spowoduje, że układ przestanie być wrażliwy na minima lokalne - będzie miał możliwość pokonania *bariery potencjału* która więzi go w minimum lokalnym. Okazuje się, że natura sama podpowiada co trzeba zrobić - odpowiedzią są fluktuacje termiczne. Wyobraźmy sobie rosnący kryształ. Ustawienie krystaliczne atomów stanowi bez wątpienia minimum energetyczne ich możliwych ustawień, jednak w naturalnych warunkach przyrost kryształu jest strasznie czuły na wszelkie mikroskopijne zaburzenia. W szczególności jeśli krystalizacja przebiega szybko, powstałe kryształki są małe a ich struktura przestrzenna jest zaburzona. Aby uzyskać duży, nieskazitelny kryształ potrzebne jest cierpliwe postępowanie polegające na powolnym obniżaniu temperatury, przy czym w momencie, gdy widzimy, że powstało jakieś zaburzenie, powinniśmy temperaturę podnieść (aby się stopiło i na nowo skrystalizowało). Taki powolny proces zwany wyżarzaniem w końcowym rezultacie powinien zapewnić duży kryształ bez żadnych zaburzeń. Temperatura okazuje się być czynnikiem, który pozwala wyjść z minimów lokalnych (niewielkich nieregularnych kryształków). Powoli obniżana powoduje, że układ jest coraz bardziej stabilny, ale ponieważ ma możliwość zwiedzenia większego obszaru przestrzeni fazowej, wybiera lepsze minimum. W dalszych rozdziałach dowiedziemy, że takie postępowanie prowadzi w granicy do znalezienia minimum globalnego z prawdopodobieństwem jeden. W informatyce opisany sposób postępowania nazywa się symulowanym wyżarzaniem (*simulated annealing*) i okazuje się przynosić spodziewane rezultaty, poprzez wprowadzenie do układu czynnika stochastycznego symulującego temperaturę. Wkraczamy tutaj na teren algorytmów stochatycznych, gdzie podstawowym narzędziem opisu ich poprawności są łańcuchy Markowa omówione szerzej w podrozdziale 3.1.

[7] Jeśli układ jest izolowany to oczywiście energia jest zachowana.

[8] W pewnych ograniczonych zagadnieniach minimalizacji jak np. minimalizacji wypukłej taka okoliczność może być częsta.

1.4. Inspiracje biologiczne w informatyce - sieci neuronowe, perceptrony, MLP

W ciągu ostatnich kilkudziesięciu lat dokonała się rewolucja informatyczna. Dzisiejsze komputery osobiste - chociaż mieszczą się w niewielkim plecaku - posiadają często moc obliczeniową która jeszcze 10 lat temu była zarezerwowana dla superkomputerów[9]. Jeśli słynne prawo Moora o podwajaniu się mocy obliczeniowej komputerów co około 18 miesięcy pozostanie prawdziwe, to za 10-20 lat moc obliczeniowa superkomputerów będzie pozwalała wykonywać obliczenia dzisiaj prawie niewyobrażalne. Jednego celu jednak nadal nie udało się osiągnąć - budowy sztucznego mózgu - inteligentnego komputera. W latach 50, gdy komputery dopiero się pojawiały wydawało się to kwestią niedalekiej przyszłości. Został nawet sformułowany test pozwalający określić czy maszyna posiada już poziom inteligencji człowieka (lub wyższy) - słynny Test Turinga - którego do tej pory nie przeszedł żaden sztuczny układ informatyczny. Pomimo, iż nadal nie można prowadzić intelektualnej konwersacji z własnym laptopem, olbrzymia ilość informacji na temat działania ludzkiego układu nerwowego została odkryta, a wiele z nich posłużyło jako inspiracje przy tworzeniu algorytmów czy też całych modeli obliczeń. Podstawowym faktem na temat budowy mózgu było to, iż składa się z neuronów - pojedynczych komórek przetwarzających sygnały elektryczne. Redukcjonistyczne podejście zakładało stworzenie przyzwoitego modelu neuronu, a następnie na zebraniu i połączeniu wielu neuronów w większą całość w nadziei stworzenia sztucznego mózgu[10]. Pionierami w tej dziedzinie byli W. S. McCulloch i W. H. Pitts którzy w swojej pracy [23] z 1943 roku podali pierwszą definicję neuronu formalnego.

Definicja 1.4.0.1 *Neuronem formalnym nazywamy układ składający się z*

- *zestawu wag* $w_1...w_n \in \{-1, 1\}$ *oraz progu* $\theta \in \mathbb{R}$.
- *zestawu aktywacji wejściowych (presynaptycznych)* $j_1...j_n \in \{0, 1\}$
- *funkcji całkowitej aktywacji:*

$$h = \sum_{i=1}^{n} w_i \cdot j_i$$

- *funkcji przejścia* ϕ

$$\phi = \begin{cases} 0 & ; h - \theta \leq 0 \\ 1 & ; h - \theta > 0 \end{cases}$$

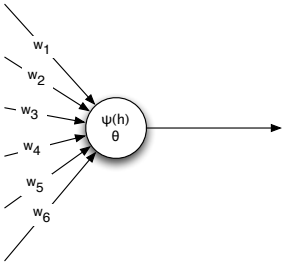
Neuron formalny zadaje zatem na przestrzeni wektorów wejściowych (binarnych, zatem rogach wielowymiarowego sześcianu) odwzorowanie do zbioru $\{0, 1\}$. Gdy przyjrzymy się bliżej temu przekształceniu okaże się, że neuron zadaje hiperpowierzchnię która tnie przestrzeń wejściową na dwie części - tych wektorów które powodują aktywację neuronu i tych które jej nie powodują. Geometryczne położenie tej płaszczyzny jest ściśle zdeterminowane przez wektor wag $w_1...w_n$ oraz próg θ. Taka a nie inna budowa neuronu jest oczywiście inspirowana biologicznie. Komórki nerwowe rzeczywiście wykazują podobne cechy - posiadają wiele wejść (dendryty) oraz jedno wyjście (akson). Sygnały przepływające z innych neuronów poprzez synapsy do dendrytów ulegają zsumowaniu, gdy dwa dendryty się łączą[11]. W efekcie

[9] To zdanie powinno być prawdziwe przynajmniej przez następne kilkanaście lat.

[10] Dzisiaj część badaczy skłania się ku twierdzeniu, iż takie redukcjonistyczne podejście nie przyniesie oczekiwanych rezultatów, istoty dynamiki mózgu trzeba raczej szukać w innych skalach.

[11] Sumowanie to w rzeczywistości ma dość skomplikowaną naturę, jest to raczej całkowanie potencjałów. Zwykła suma daje jednak zadowalające przybliżenia, a ma tę zaletę że jest łatwa do policzenia.

Tylko do użytku edukacyjnego



Rysunek 1.1. Schemat Neuronu formalnego

Źródło: Opracowanie własne

do ciała neuronu dociera zsumowany sygnał ze wszystkich dendrytów. Jeśli potencjał ten jest wstanie wzbudzić neuron, wysyła on silny sygnał do aksonu, w przeciwnym wypadku sygnał zamiera (temu zachowaniu odpowiada funkcja przejścia). Zdolność do przewodzenia poszczególnych synaps która w rzeczywistości zależy od kształtu synapsy i składu neuroprzekaźników w modelu McCullocha-Pittsa jest zastąpiona poprzez zestaw wag. Widać zatem, że pomimo znacznej prostoty neuron formalny odpowiada w miarę dobrze rzeczywistości.

Praca McCullocha-Pittsa spowodowała pewne zamieszanie i doprowadziła do reanimacji idei sztucznego mózgu. W istocie była inspiracją dla wielu badaczy, między innymi dla Kleenego, który w toku rozważań nad sieciami neuronowymi wymyślił idee automatu skończonego. Na bazie neuronu formalnego można także skonstruować dowolną bramkę realizującą monotoniczną funkcję logiczną.

Pewnym ograniczeniem idei neuronu formalnego jest fakt, że pracuje on jedynie na danych binarnych. W dalszych pracach rozwinięto tę ideę do modelu tak zwanego *perceptronu*[12], w którym wagi mogą być dowolnymi liczbami rzeczywistymi. Powoduje to, że płaszczyzny separujące mogą być ustawione w przestrzeni pod dowolnymi kątami, daje to więcej swobody i umożliwia automatyczne uczenie[13]. Rozpoczęto także eksperymenty z różniczkowalnymi funkcjami aktywacji, co umożliwiało stosowanie algorytmów spadku gradientowego przy uczeniu.

Na początku lat 60 wydawało się, że teraz wystarczy jedynie skonstruować układ składający się z wielu neuronów i że przejście testu Turinga jest bliskie. Szybko okazało się jednak, iż możliwości ówczesnych modeli neuronowych nie są aż tak duże - pojedynczy neuron potrafi rozwiązywać jedynie zagadnienia liniowo separowalne (czyli takie w których zbiór rozwiązań problemu od reszty daje się oddzielić hiperpłaszczyzną), klasa tych problemów okazała się dość skromna, wystarczy nadmienić, iż problem XOR (binarnej różnicy symetrycznej) nie jest liniowo separowalny. Aby poszerzyć klasę problemów rozwiązywanych przez sieci neuronowe, trzeba było konstruować sieci wielowarstwowe. Jak się jednak okazało ówczesne algorytmy uczenia nie były przewidziane dla układów wielowarstwowych i nie było zadowalających rozwiązań tego problemu. Po serii porażek idea sieci neuronowych została uśpiona na dobre 20 lat.

[12]Perceptron to dość ogólne pojęcie związane z modelowaniem zjawiska percepcji czyli rozpoznawania cech obrazów. Perceptrony były zaproponowane w 1958 roku przez Franka Rossenblatta a następnie badane przez Paperta i Minskiego

[13]Można pokazać, że moc obliczeniowa sieci ważonych i nieważonych jest taka sama, jednak struktura topologiczna tych drugich może być bardzo skomlikowana (redundantne synapsy).

Kolejny renesans sieci neuronowe przeżyły w połowie lat 80 dwudziestego wieku, po publikacji Rumelharta, Hintona i Williamsa [34] prezentującej elegancki i prosty w implementacji algorytm wstecznej propagacji błędu[14] (Backpropagation algorithm) pozwalający uczyć wielowarstwowe sieci MLP[15] składające się z perceptronów. Algorytm ten jest typowym przykładem algorytmu spadku gradientowego, w wyniku jego działania minimalizowany jest błąd średniokwadratowy rzeczywistego wyjścia z sieci względem oczekiwanego wyjścia. Elegancja tego rozwiązania polegała na tym, że stosując funkcje sigmoidalne wykładnicze, nie trzeba w czasie działania algorytmu liczyć żadnych pochodnych - liczą się w sprytny sposób same.

Od tego czasu do dnia dzisiejszego z mniejszym lub większym natężeniem wraca moda na sieci neuronowe, pojawiają się coraz to nowe idee (sieci Hopfielda, maszyny Boltzmanna, sieci Bayesowskie[16], sieci samoorganizujące Kohonena), oraz ciekawe zastosowania. Sztuczne sieci neuronowe są tak naprawdę silnie obecne w naszym codziennym życiu, we wszelkiego rodzaju zaawansowanych czujnikach, cyfrowych aparatach fotograficznych a także w zastosowaniach militarnych, celownikach rakiet itp. Szczegóły dotyczące tematu sieci neuronowych można znaleźć w jednej z licznych monografii na ten temat [33, 10, 27, 15].

1.5. Przegląd algorytmów minimalizacji globalnej

Wróćmy na chwilę do tematu minimalizacji globalnej. Zgodnie z argumentacją zawartą w 1.1, praktycznie każdy problem informatyczny daje się sformułować w języku minimalizacji globalnej, stąd duże zainteresowanie tym zagadnieniem. Powstał szereg algorytmów minimalizacji globalnej które w praktyce wszystkie sprowadzają się do przeszukania całej dziedziny minimalizowanej funkcji (w ogólności nie wygląda by dało się tego uniknąć[17]). W zależności od charakteru problemów różne strategie przeglądania dziedziny problemu mogą przynieść różne rezultaty, głównie w kluczowej kwestii czasu działania algorytmu.

1.5.1. Algorytm Random Walk

Najprostszy stochastyczny algorytm poszukiwania minimum globalnego, polegający na losowym błądzeniu po przestrzeni rozwiązań. Zależnie od wersji, kolejny stan wybiera się w losowym punkcie przestrzeni, lub jako jeden z sąsiadujących punktów obecnego stanu (wtedy losuje się kierunek przemieszczania.

1.5.2. Algorytm Tabu-Search

[18] Pewne modyfikacja algorytmu Random Walk, polegająca na tym, że raz odwiedzone stany oznacza się jako zakazane (tabu). Chroni to algorytm przed bezsensownym zwiedzaniem raz już spenetrowanych fragmentów przestrzeni rozwiązań (i powiększa prawdopodobieństwo odwiedzenia bardziej oddalonych obszarów), jednak wymaga pamiętania stanów które już odwiedzono (co może okazać się kłopotliwe, gdy stanów będzie dużo[19]).

[14] Algorytm ten równolegle opisany był przez Davida Parkera w pracy [25], jednak po raz pierwszy został przedstawiony przez Paula Werbosa w pracy doktorskiej obronionej na Harvardzie 1974 [36] a następnie zapomniany na kilka lat.

[15] MLP -Multi Layer Perceptron - Perceptron wielowarstwowy

[16] Chociaż sieci Bayesowskie nie są sieciami neuronowymi, pojawiają się jednak w szerszej dziedzinie inteligencji obliczeniowej, w której właśnie sieci neuronowe wiodły prym.

[17] Algorytm minimalizacji globalnej, który zawsze przeszukiwałby jedynie logarytmiczną część przestrzeni zapewniałby rozwiązanie wartego milion dolarów słynnego problemu informatycznego $P = NP$.

[18] Zarówno pisownia Taboo jak i Tabu jest w użyciu

[19] W praktyce pamięta się tylko pewną ilość odwiedzonych stanów. Są też wersje tabu-search które dynamicznie modyfikują ilość pamiętanych stanów (Reactive tabu search)

1.5.3. Algorytm wielu startów (multistart)

W algorytmie tym (tak naprawdę pewnej strategii postępowania) wykorzystuję się dowolny algorytm minimalizacji lokalnej (np. algorytm spadku gradientowego), a następnie uruchamia go wiele razy z losowo wybranych punktów startowych. Przy odrobinie szczęścia można natrafić na punkt startowy leżący w basenie atrakcji minimum globalnego, wtedy algorytm minimalizacji lokalnej niechybnie do niego trafi. Niestety baseny atrakcji minimów globalnych mogą zajmować tylko bardzo nieznaczny fragment przestrzeni fazowej, a w dodatku mogą mieć bardzo skomplikowaną geometrię co dodatkowo utrudnia trafienie w stosowny punkt.

1.5.4. Algorytm P*

Jest to dość ciekawy algorytm oparty o strategię *Random Line Search*[20]. Polega on na wylosowaniu kierunku (z rozkładu jednostajnego) oraz długości odcinka (z rozkładu normalnego). Następnie na wylosowanym odcinku stosowany jest jednowymiarowy algorytm minimalizacji (może być to po prostu losowe próbkowanie odcinka w poszukiwaniu najlepszego minimum). Znalezione minimum jest obierane jako kolejny punkt startowy. Losowanie długości przeszukującego odcinka z rozkładu normalnego powoduje, że większość poszukiwań ma charakter lokalny, jednak zawsze jest możliwość wyjścia z minimum lokalnego.

1.5.5. Simulated Annealing

Algorytm którego aspektami będziemy się zajmować w dalszych fragmentach tej pracy. Wykorzystuje on ciekawe własności pewnej rodziny rozkładów prawdopodobieństwa na przestrzeniach fazowych (miar Gibbsa), oraz fakt, że losując stan z tych rozkładów a jednocześnie przechodząc z temperaturą układu do 0 trafia się w minimum globalne z prawdopodobieństwem 1. Szczegóły dotyczące tej rodziny algorytmów przedstawione są w następnych rozdziałach.

Obszerny przegląd dostępnych algorytmów minimalizacji globalnej (jest ich znacznie więcej niż te wymienione wyżej) można znaleźć w pracy [9].

[20] Przeszukiwanie losowymi liniami

Tylko do edukacyjnych

ROZDZIAŁ 2

Sieci rekurencyjne, sieci Hopfielda

Na początku lat 80 XX wieku, w czasie kiedy tradycyjne warstwowe sieci neuronowe oparte na perceptronach (sieci jednokierunkowe, tzw *feed forward networks*) przeżywały kryzys związany z brakiem algorytmu uczącego, pojawiła się nowe idea, która jak się później okazało była wręcz rewolucyjna i niezwykle owocna naukowo. Całe zamieszanie zapoczątkowała opublikowana w 1982 roku praca J.J. Hopfielda [19], o wiele mówiącym tytule "Neural networks and physical systems with emergent collective computational abilities" [1]. W pracy tej nie tylko był opisany model obliczeń którego podstawę stanowiła rekurencyjna sieć neuronowa, ale także wskazywał algorytm uczenia (obliczenia wag) i dynamikę prowadzącą do rozwiązania problemu autoasocjacji. Opublikowanie tych wyników spowodowało duże zainteresowanie świata naukowego które w szybkim czasie zaowocowało setkami prac analizującymi sieci Hopfielda (sama oryginalna praca Hopfielda była do dzisiaj cytowana co najmniej 2000 razy) a także pokaźnym zbiorem implementacji sprzętowych i zastosowań.

2.1. Definicja, przykłady dynamiki

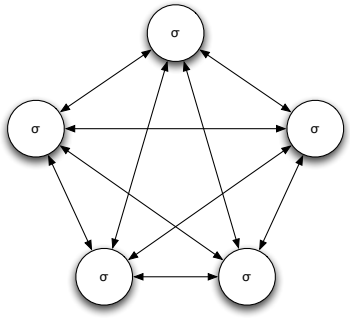
Definicja 2.1.0.1 *Siecią rekurencyjną będziemy nazywać układ składający się z*

- *neuronów (spinów)* $\sigma_i \in \{0, 1\}$*, przy czym* $i \in 1..n$
- *zestawu wag* $w_{i,j} \in \mathbb{R}$*, przy czym* $w_{i,j}$ *oznacza wagę połączenia synaptycznego z neuronu* σ_i *do neuronu* σ_j
- *wektora* $h \in \mathbb{R}^n$ *zwanego dalej wektorem pola zewnętrznego*
- $\tau_p \in \mathbb{R}$ *- czasu przejścia impulsu po połączeniu*
- $\tau_r \in \mathbb{R}$ *- czasu refrakcji neuronu (minimalnego czasu pomiędzy dwoma aktywacjami)*

Na początku warto zauważyć, że sieć rekurencyjna jest całkowicie sprzężona, w przeciwieństwie do rozważanych wcześniej sieci MLP, w których sprzężenia zwrotne wszakże czasem występowały, jednak były w wielu kwestiach kłopotliwe i utrudniały opis teoretyczny oraz uczenie takich sieci. W przypadku sieci rekurencyjnych sprzężenie jest całkowite i zupełnie pożądane. Czasy refrakcji i czasy przejścia impulsu, pochodzą z rozważań biologicznych. W rzeczywistych sieciach neuronowych występują opóźniania w propagacji sygnałów wynikające mechanizmu działania pompy jonowej w błonie neuronu. Po aktywacji neuron także przez pewien czas nie wykazuje gotowości do aktywacji (czas refrakcji). Dokładniej zjawiska występujące w biologicznych sieciach neuronowych modelują tak zwane sieci impulsujące opisane w [3]. Sieci te mają dość skomplikowaną dynamikę czasu ciągłego opartą na impulsach propagujących się z pewnymi opóźnieniami i mocami w synapsach. Gdy aktywacja łączna neuronu przekroczy pewien próg, wyzwalany jest sygnał, a następnie neuron jest usypiany na pewien czas (refrakcja). W modelu sieci rekurencyjnych nie rozważamy ewolucji w czasie ciągłym, parametry τ_p i τ_r mają jedynie znaczenie przy obieraniu globalnej dynamiki dla sieci.

[1] Sieci neuronowe i układy fizyczne ze spontanicznymi zbiorowymi możliwościami obliczeniowymi

Tylko do celów edukacyjnych



Rysunek 2.1. Schemat pięcioelementowej sieci rekurencyjnej

Źródło: Opracowanie własne

- gdy $\tau_p < \tau_r$ mamy dynamikę asynchroniczną (tzw. dynamikę Glaubera) w której
  - losujemy neuron σ_i
  - ustawiamy go w kierunku pola wypadkowego Σ_i (o ile jest ono niezerowe) (będącego sumą pola lokalnego i pola zewnętrznego)

    $$\sigma_i = sgn\left(\Sigma_i\right) = sgn\left(\sum_{j=1}^{n} w_{j,i} + h_i\right)$$

    Założenia braku modyfikacji przy zerowym polu będzie wygodne w pewnych zbliżających się dowodach.
- gdy $\tau_p \sim \tau_r$ mamy dynamikę synchroniczną (tzw. dynamikę Little'a w której stan wszystkich neuronów zmienia się naraz.
- gdy $\tau_r < \tau_p$ mamy dynamikę hybrydową, gdzie sieć jest podzielona na funkcjonalne fragmenty w których obowiązuje dynamika Glaubera, natomiast fragmenty pomiędzy sobą synchronizują się dynamiką synchroniczną (przesyłając średnią aktywację jako sygnał).

W naszych dalszych rozważaniach będziemy zakładać, że układ pracuje w dynamice asynchronicznej, zaś macierz wag $w_{i,j}$ jest symetryczna. Założenia te wynikają z konieczności uproszczenia układu do analiz teoretycznych (dynamika asynchroniczna posiada stan końcowy dla prawie każdego układu wag[2] i nie popada w cykle, zaś założenie symetryczności macierzy $w_{i,j}$ znakomicie ułatwia operowanie wzorami na energię układu). Zakładamy też brak samopołączeń, czyli $w_{i,i} = 0$.

Definicja 2.1.0.2 *Energią konfiguracji $\overline{\sigma}$[3] sieci nazywamy funkcję*

$$E(\overline{\sigma}) := -\frac{1}{2}\sum_{i,j} w_{i,j}\sigma_i\sigma_j - \sum_i h_i\sigma_i$$

[2]Można na siłę wygenerować patologiczne wagi które wprowadzą degenerację, takich układów wag jest jednak mało (miary zero).

[3]Przez $\overline{\sigma}$ będziemy oznaczać pojedynczy stan sieci (konfigurację wszystkich neuronów)

Tylko do celów edukacyjnych

Kluczowe znaczenie ma w tym miejscu następujące

Twierdzenie 2.1.0.1 *Energia konfiguracji sieci nie wzrasta w czasie ewolucji w dynamice asynchronicznej.*

Dowód 2.1.0.2 *Oznaczmy przez $\overline{\sigma}$ stan wyjściowy sieci, a przez $\overline{\sigma}'$ stan sieci po dokonaniu zmiany przez dynamikę. Załóżmy, że zmieniony został neuron i (w przypadku, gdy nie został zmieniony żaden neuron, twierdzenie jest oczywiste).*

$$\begin{aligned} E(\overline{\sigma}') - E(\overline{\sigma}) &= -\frac{1}{2}\sum_j w_{i,j}\sigma_j(\sigma_i' - \sigma_i) - h_i(\sigma_i' - \sigma i) \\ &= -(\sigma_i' - \sigma_i)\Sigma_i \leq 0 \end{aligned}$$

Ponadto jeśli założymy, że przy zerowym polu lokalnym nigdy nie następują zmiany to powyższa nierówność jest ostra. Ponieważ energia jest ograniczona z dołu (z góry też) to oznacza to, że w dynamice synchronicznej osiągany jest zawsze pewien atraktor.

□

Udowodnione twierdzenia pokazują, że sieć rekurencyjna (w dynamice asynchronicznej) z dowolnego losowego stanu początkowego zawsze zbiegnie do minimum lokalnego, gdzie dynamika się zatrzymuje. W ogólności nie musi to być jednak jedno z oczekiwanych minimów, gdyż w trakcie ustalenia wag mogą powstać tak zwane *atraktory pasożytnicze*, czyli pewne dodatkowe minima lokalne funkcji energetycznej (problem ten będzie jaśniej opisany niżej na przykładzie sieci zapamiętującej wzorce).

2.2. Sieci Hopfielda, pojemność pamięciowa, katastrofa zapominania

Sieci rekurencyjne często stosuje się jako autoasocjator. *Autoasocjator*, to rodzaj pamięci skojarzeniowej [4] pamiętającej szereg wzorców, który umożliwia odtworzenie każdego z nich na podstawie zaszumionego przykładu. W tym kontekście sieci rekurencyjne nazywane są sieciami Hopfielda, który jako pierwszy w cytowanej już wyżej pracy [19] opisał takie zastosowania. Interesujące jest pytanie o maksymalną pojemność pamięciową sieci (maksymalną ilość wzorców które sieć potrafi efektywnie zapamiętać, a następnie poprawnie odtworzyć), a także o zjawiska które mogą się pojawić przy ewentualnym przeuczeniu. W dalszej części tego rozdziału będziemy zakładać:

- Badana sieć składa się z N neuronów, gdzie N jest w miarę duże[5]
- Mamy danych P wzorców losowych niezależnych i nieobciążonych ($\mathbb{E}(\xi_i) = 0$)

$$\mathbb{I}^\mu = \{\xi_i\}^\mu, \quad i = 1...N$$

- Definiujemy stopień zgodności $\mathbb{M}^\mu(\overline{\sigma})$ stanu $\overline{\sigma}$ ze wzorcem $\mathbb{I}^\mu$

$$\mathbb{M}^\mu(\overline{\sigma}) = \frac{1}{N}\sum_i \sigma_i \xi_i^\mu = \frac{1}{N}\langle \overline{\sigma}, \mathbb{I}^\mu \rangle$$

[4] Pamięć asocjacyjna (skojarzeniowa) - układ, który pobudzony pewnymi danymi zwraca skojarzone z nimi inne dane. Często mówi się też o pamięci adresowanej danymi.

[5] Tak duże, żeby przyzwoicie działało centralne twierdzenie graniczne.

Tylko do celów edukacyjnych

- Definiujemy funkcję energetyczną (z dokładnością do stałej) jako średniokwadratowy błąd odtworzenia wzorców:

$$E(\overline{\sigma}) = -\frac{1}{2}N\sum_{\mu}\left(\mathbb{M}^{\mu}(\overline{\sigma})\right)^2$$

- Obserwujemy symetryczną degeneracje sieci która powoduje, że w obliczu braku pola zewnętrznego stan $\overline{\sigma}$ zachowuje się identycznie jak $-\overline{\sigma}$

- Ustalamy wagi dopasowując je do postaci funkcji energetycznej jak następuje

$$\begin{aligned} E(\overline{\sigma}) &= \frac{-1}{2N}\sum_{\mu}\sum_{i,j}\sigma_i\xi_i^{\mu}\xi_j^{\mu}\sigma_j - \frac{1}{2}P \equiv \\ &\equiv \frac{-1}{2N}\sum_{\mu}\sum_{i,j}\sigma_i\xi_i^{\mu}\xi_j^{\mu}\sigma_j = \\ &= \frac{-1}{2}\sum_{i,j}\sigma_i\sigma_j w_{i,j} \end{aligned}$$

  stąd wagi $w_{i,j}$ zadają się wzorem

$$w_{i,j} = \frac{1}{N}\sum_{\mu}\xi_i^{\mu}\xi_j^{\mu} \tag{2.1}$$

  Wagi ustalone w ten sposób nawiązują do reguły Hebba zapoponowanej w [16] polegającej na wzmacnianiu połączeń między neuronami o takich samych stanach.

Naszą analizę pojemności pamięciowych sieci Hopfielda[6] rozpoczniemy od zbadania stabilności pojedynczego wzorca. Niech stan sieci $\overline{\sigma} = \mathbb{I}^{\mu_0}$. Pokażemy że stan $\mathbb{I}^{\mu_0}$ jest punktem stałym dynamiki asynchronicznej. Wybierzmy neuron i. Pole lokalne dla tego neuronu rozpisuje się jako:

$$\begin{aligned} \Sigma_i(\overline{\sigma}) &= \sum_j w_{i,j}\sigma_j = \sum_j\left(\frac{1}{N}\sum_{\mu}\xi_i^{\mu}\xi_j^{\mu}\right)\sigma_j = \\ &= \sum_{\mu}\mathbb{M}^{\mu}(\overline{\sigma})\xi_i^{\mu} \end{aligned}$$

Ostatecznie, gdy $\overline{\sigma} = \mathbb{I}^{\mu_0}$ pole ma postać

$$\Sigma_i(\mathbb{I}^{\mu_0}) = \xi_i^{\mu_0} + \sum_{\mu\neq\mu_0}\mathbb{M}^{\mu}(\mathbb{I}^{\mu_0})\xi_i^{\mu}$$

Dwa czynniki w powyższej sumie określa się jako impuls ($\xi_i^{\mu_0}$) i szum ($\sum_{\mu\neq\mu_0}\mathbb{M}^{\mu}(\mathbb{I}^{\mu_0})\xi_i^{\mu}$) Zatem wzorzec $\mathbb{I}^{\mu_0}$ jest stabilny w i o ile znak pola lokalnego pokrywa się ze znakiem wzorca, co po rozważeniu dwóch przypadków implikuje

$$\xi_i^{\mu_0}\left(\sum_{\mu\neq\mu_0}\mathbb{M}^{\mu}(\mathbb{I}^{\mu_0})\xi_i^{\mu}\right) > -1$$

[6] Jest to dość klasyczne rozumowanie opisane po części juz w oryginalnej pracy [19] Hopfielda.

Tylko do celów edukacyjnych

Jednym z naszych założeń było, iż wzorce są wylosowane niezależnie, ($\xi_i^\mu = +1$ z prawdopodobieństwem $p = \frac{1}{2}$ oraz $\xi_i^\mu = -1$ z prawdopodobieństwem $p = \frac{1}{2}$). Ponieważ N jest odpowiednio duże zatem

$$\mathbb{M}^\mu(\mathbb{I}^{\mu_0}) \sim \mathcal{N}\left(0, \frac{1}{N}\right)$$

oraz konsekwentnie

$$\sum_{\mu \neq \mu_0} \mathbb{M}^\mu(\mathbb{I}^{\mu_0})\xi_i^\mu \sim \mathcal{N}\left(0, \frac{P-1}{N}\right)$$

Z czego daje się wywnioskować, że musi zachodzić $P \ll N+1$ aby szum nie stanowił groźnej konkurencji dla czynnika impulsowego. Rozważmy teraz możliwości usuwania zaburzeń przez sieć Hopfielda. Niech $\hat{\mathbb{I}}$ oznacza wzorzec $\mathbb{I}^{\mu_0}$ zaburzony w R miejscach.Wiadomo, że

$$\Sigma_i(\overline{\sigma}) = \sum_\mu \mathbb{M}^\mu\left(\overline{\sigma}\right)\xi_i^\mu$$

oraz

$$\begin{aligned}\Sigma_i\left(\mathbb{I}^{\mu_0}\right) &= \sum_{\mu \neq \mu_0} \mathbb{M}^\mu\left(\mathbb{I}^{\mu_0}\right)\xi_i^\mu + \xi_i^{\mu_0}\mathbb{M}^{\mu_0}\left(\mathbb{I}^{\mu_0}\right) = \\ &= \sum_{\mu \neq \mu_0} \mathbb{M}^\mu\left(\mathbb{I}^{\mu_0}\right)\xi_i^\mu + \xi_i^{\mu_0} \cdot 1\end{aligned}$$

zatem

$$\begin{aligned}\Sigma_i(\hat{\mathbb{I}}) &= \sum_{\mu \neq \mu_0} \mathbb{M}^\mu\left(\hat{\mathbb{I}}\right)\xi_i^\mu + \xi_i^{\mu_0}\mathbb{M}^{\mu_0}\left(\hat{\mathbb{I}}\right) = \\ &= \sum_{\mu \neq \mu_0} \mathbb{M}^\mu\left(\hat{\mathbb{I}}\right)\xi_i^\mu + \xi_i^{\mu_0} \cdot \left(1 - \frac{2R}{N}\right)\end{aligned}$$

przy czym (centralne twierdzenie graniczne)

$$\sum_{\mu \neq \mu_0} \mathbb{M}^\mu\left(\hat{\mathbb{I}}\right)\xi_i^\mu \sim \mathcal{N}(0, \alpha)$$

Niech $\alpha = \frac{P}{N}$ oraz $\alpha << 1$. Wtedy prawdopodobieństwo że i-ty neuron zostanie poprawnie odtworzony, po przybliżeniu przez centralne twierdzenie graniczne, wynosi

$$\begin{aligned}&\mathbb{P}\left(\left[\mathcal{N}(0,\alpha) + \xi_i^{\mu_0}\left(1 - \frac{2R}{N}\right)\right]\xi_i^{\mu_0} > 0\right) = \\ =&\mathbb{P}\left(\mathcal{N}(0,\alpha) < 1 - \frac{2R}{N}\right) = \\ =&\Phi\left(\frac{1 - \frac{2R}{N}}{\sqrt{\alpha}}\right) = \\ =&1 - \overline{\Phi}\left(\frac{1 - \frac{2R}{N}}{\sqrt{\alpha}}\right)\end{aligned}$$

Prawdopodobieństwo tego, że stany wszystkich neuronów zostaną poprawnie odtworzone można oszacować jako

$$\left(1 - \overline{\Phi}\left(\frac{1 - \frac{2R}{N}}{\sqrt{\alpha}}\right)\right)^N$$

Tylko do celów edukacyjnych

co po zastosowaniu przybliżenia pierwszego rzędu można oszacować przez

$$1 - N\overline{\Phi}\left(\frac{1-\frac{2R}{N}}{\sqrt{\alpha}}\right)$$

W dalszych rozważaniach będziemy korzystać z następującego przybliżenia

$$\overline{\Phi}(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{s^2}{2}}ds \sim \frac{1}{\sqrt{2\pi}x}e^{-\frac{x^2}{2}}$$

Oszacowanie to jest poprawne ze względu na wartość granicy

$$\lim_{x\to\infty}\frac{\frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{s^2}{2}}ds}{\frac{1}{\sqrt{2\pi}x}e^{-\frac{x^2}{2}}} = 1$$

Przyjmijmy teraz, że dopuszczamy prawdopodobieństwo błędu na poziomie δ. Wtedy

$$\begin{aligned}\delta > & N\overline{\Phi}\left(\frac{1-\frac{2R}{N}}{\sqrt{\alpha}}\right) \sim \\ & \sim N\frac{1}{\sqrt{2\pi}}\sqrt{\alpha}\frac{1}{1-\frac{2R}{N}}e^{\left(\frac{-\left(1-\frac{2R}{N}\right)^2}{2\alpha}\right)} \sim \\ & \sim \frac{N\sqrt{\alpha}}{\sqrt{2\pi}}e^{\left(\frac{-\left(1-\frac{2R}{N}\right)^2}{2\alpha}\right)}\end{aligned}$$

z czego po dalszych przekształceniach otrzymujemy

$$\begin{aligned}\ln\left(\delta\frac{\sqrt{2\pi}}{N\sqrt{\alpha}}\right) &> -\frac{\left(1-\frac{2R}{N}\right)^2}{2\alpha} \\ 2\alpha\ln\left(\delta\frac{\sqrt{2\pi}}{N\sqrt{\alpha}}\right) &> -\left(1-\frac{2R}{N}\right)^2 \\ 2\alpha\ln\left(\frac{N\sqrt{\alpha}}{\sqrt{2\pi}\delta}\right) &> \left(1-\frac{2R}{N}\right)^2\end{aligned}$$

stąd

$$\alpha \sim \frac{\left(1-\frac{2R}{N}\right)^2}{2\log(N)}$$

Po wzięciu pod uwagę faktu, że przestrzeń fazowa stanów sieci Hopfielda jest symetryczna ze względu na branie negatywu stanu, otrzymany wynik trzeba podzielić przez 2, co ostatecznie daje

$$\alpha \sim \frac{\left(1-\frac{2R}{N}\right)^2}{4\log(N)}$$

z czego wnioskujemy, że

$$P \sim N\frac{\left(1-\frac{2R}{N}\right)^2}{4\log(N)}$$

Tylko do celów edukacyjnych

Ponieważ R jest z założenia znacznie mniejsze niż N, zatem

$$P \sim \frac{N}{4\log(N)}$$

Jest to dość zgrubne (jednak nadal przyzwoite) oszacowanie pojemności pamięciowej dla N neuronowej sieci Hopfielda. Na tej podstawie pewnych innych oszacowań (formalnie mających status hipotez) oraz wyników eksperymentalnych [21] można określić pojemność pamięciową sieci przy $N \to \infty$ na $\mathfrak{a} = 0,138$ (oznacza to, że $\lim_{N\to\infty} \frac{\max_N P}{N} \sim 0,138$). Przy większym nasyceniu wzorcami występuje zjawisko zwane *katastrofą zapominania*. Zjawisko to polega na tym, że sieć nagle przestaje pamiętać praktycznie wszystkie zapamiętane wcześniej wzorce, przestają one być punktami stabilności dynamiki, dominującą rolę zaczynają natomiast odgrywać atraktory pasożytnicze. Siec po przekroczeniu granicy pojemności staje się zupełnie bezużyteczna.

2.3. Przykłady problemów przetłumaczonych na język sieci rekurencyjnych

Jak się okazuje, sieci rekurencyjne można traktować jako uniwersalne narzędzie obliczeniowe, które przy pomocy odpowiedniego kodowania i odpowiedniego dobrania funkcji energetycznej może rozwiązywać różnorodne problemy informatyczne. Pewną cechą wspólną przedstawionych poniżej przykładów problemów rozwiązywanych przez sieć rekurencyjne jest rozdział funkcji energetycznej na dwa człony:

- E_{ogr} - energia ograniczeń - człon, który minimalizuje się jeśli stan sieci odpowiada dopuszczalnemu rozwiązaniu problemu, osiąga dużą wartość, gdy stan sieci odpowiada niedopuszczalnemu (niepoprawnemu) rozwiązaniu problemu.[7]
- E_{opt} - energia optymalizacji - człon odpowiadający rzeczywistej optymalizacji problemu.

Sieć w toku ewolucji powinna być karana za wszelkie niepoprawne rozwiązania (wzrost energii ograniczeń), gdy już znajdzie się w poprawnym rozwiązaniu powinna je optymalizować (minimalizacji energii optymalizacji). Przedstawione poniżej problemy są klasycznymi przykładami problemów *NP-zupełnych*. Ponieważ wszystkie takie problemy są ze sobą równoważne, zatem jedynie poprzez zmianę kodowania sieć rekurencyjna może rozwiązywać dowolny problem NP-zupełny.[8]

2.3.1. Problem kolorowania grafu

Dany jest graf planarny G o N wierzchołkach. Szukamy takiego pokolorowania grafu na 4 różne kolory, by żadne dwa sąsiednie wierzchołki nie były tego samego koloru.

- Kodowanie - niech $v_{i,j}$ będzie macierzą sąsiedztwa grafu G. Bierzemy $4 \cdot N$ neuronów $\sigma_{i,\alpha}$, przy czym $i = 1...N$ oraz $\alpha = 1...4$. Stan neuronu definiujemy następująco

$$\sigma_{i,\alpha} = \begin{cases} 1 & ;\text{jeśli i-ty wierzchołek grafu ma kolor } \alpha \\ 0 & ;\text{w przeciwnym przypadku} \end{cases}$$

[7] Na przykład, gdy poszukujemy najkrótszego cyklu Hamiltona, a sieć koduje rozwiązanie które rzeczywiście jest dość krótkie, ale wcale nie jest cyklem!

[8] Sieci rekurencyjne nie są jednak lekarstwem na wszelkie problemy obliczeniowe. W rzeczywistości symulacja rozwiązania danego problemu przez sieć na klasycznym komputerze zajmuje zazwyczaj znacznie więcej czasu niż realizacja klasycznego algorytmu (nawet brute-force) lub heurystyki. Zyski mogą się jednak pojawić jeśli sieci realizuje się poprzez dedykowane równoległe układy elektroniczne.

Tylko do celów edukacyjnych

- Energia ograniczeń

$$E_{ogr}(\overline{\sigma}) = \lambda_0 \sum_i \left(\sum_\alpha \sigma_{i,\alpha} - 1 \right)^2 \equiv \quad \text{(równość z dokładnością do stałej)}$$
$$\equiv \lambda_0 \sum_i \left[\sum_\alpha \sum_{\beta \neq \alpha} \sigma_{i,\alpha}\sigma_{i,\beta} - \sum_\alpha \sigma_{i,\alpha} \right]$$

  Zauważmy, że energia ograniczeń rzeczywiście minimalizuje się dokładnie wtedy, gdy każdy wierzchołek grafu ma w kodowaniu neuronowym przypisany dokładnie jeden kolor.

- Energia optymalizacji

$$E_{opt}(\overline{\sigma}) = \lambda_1 \sum_i \sum_j v_{i,j} \sum_\alpha \sigma_{i,\alpha}\sigma_{j,\alpha}$$

  Zauważmy, że energia optymalizacji osiąga minimum w sytuacji w której każde dwa połączone ze sobą wierzchołki grafu są różnego koloru.

Ostatecznie energia konfiguracji wyraża się następująco

$$E(\overline{\sigma}) = \sum_{i,\alpha,j,\beta} \left(\lambda_0 \delta_{i,j}(1 - \delta_{\alpha,\beta}) + \lambda_1 v_{i,j}\delta_{\alpha,\beta} \right) \sigma_{i,\alpha}\sigma_{j,\beta} - \lambda_0 \sum_{i,\alpha} \sigma_{i,\alpha}$$

przy czym $\delta_{i,j}$ jest standardową deltą Kroneckera natomiast λ_0 i λ_1 są parametrami rzeczywistymi wyznaczającymi równowagę między członami energetycznymi[9]. Wagi oraz pole zewnętrzne określamy jak poniżej

$$w_{(i,\alpha),(j,\beta)} = -2 \left(\lambda_0 \delta_{i,j}(1 - \delta\alpha, \beta) + \lambda_1 v_{i,j}\delta_{\alpha,\beta} \right)$$
$$h_{i,\alpha} = \lambda_0$$

2.3.2. Problem planu zajęć

Danych jest N_h godzin, N_p nauczycieli oraz N_c klas. Nauczyciel p uczy w klasie c przez $n_{p,c}$ godzin, przy czym zakładamy że $\sum_c n_{p,c} \leq N_h$.

- Kodowanie - rozważmy $N_h \cdot N_p \cdot N_c$ jednostek kodujących plany zajęć następująco:

$$\sigma_{h,p,c} = \begin{cases} 1 & ;\text{jeśli w godzinie } h \text{ nauczyciel } p \text{ uczy klasę } c \\ 0 & ;\text{w przeciwnym przypadku} \end{cases}$$

[9]Staranne dobranie tych parametrów ma kluczowe znaczenie dla działania sieci. Jeśli bowiem energia ograniczeń będzie miała zbyt duży wkład, sieć osiągnie zadowalające ją minimum w stanie dopuszczalnym bez wysilania się na optymalizację. Jeśli natomiast wkład energii ograniczeń będzie zbyt mały, sieć osiągnie stan o bardzo niskiej energii optymalizacji, jednak prawdopodobnie w niedopuszczalny (jeden wierzchołek będzie miał przypisany wiele kolorów itp.).

Tylko do celów edukacyjnych

- Energia ograniczeń

$$E_{ogr} = \lambda_1 \sum_{p,c} \left(\sum_h \sigma_{h,p,c} - N_{p,c} \right)^2 + \\ + \lambda_2 \sum_{h,c} \left(\sum_p \sigma_{h,p,c} \right) \left(\left(\sum_p \sigma_{hpc} \right) - 1 \right) + \\ + \lambda_2 \sum_{h,p} \left(\sum_c \sigma_{h,p,c} \right) \left(\left(\sum_c \sigma_{hpc} \right) - 1 \right)$$

  Przy tak zadanej energii ograniczeń[10] wagi i pole zewnętrzne wyrażają się

$$w_{(h,p,c),(h',p',c')} = -2\lambda_1 \delta_{p,p'} \delta_{c,c'} (1 - \delta_{h,h'}) - 2\delta_{h,h'} [\lambda_2 \delta_{c,c'} (1 - \delta_{p.p'}) + \lambda_3 \delta_{p,p'} (1 - \delta_{c,c'})] \\ h_{h,p,c} = -\lambda_1 (1 - 2N_{pc})$$

- Energia optymalizacji. Optymalizacja zadania ułożenia planu zajęć może mieć różny charakter, tutaj dla przykładu założymy, że chcielibyśmy aby zajęcia odbywały się równomiernie w czasie (aby średnia ilość zajęć na godzinę była w miarę stała). Wprowadzamy średnią ilość zajęć na godzinę

$$\overline{N} = \frac{\sum_{p,c} N_{p,c}}{N_h}$$

  i definiujemy energię optymalizacji jako

$$E_{opt} = \lambda_0 \sum_h \left(\sum_{p,c} \sigma_{h,p,c} - \overline{N} \right)$$

co owocuje następującymi wagami i polem zewnętrznym

$$w_{(h,p,c),(h',p',c')} = -2\lambda_1 \delta_{p,p'} \delta_{c,c'} (1 - \delta_{h,h'}) - 2\delta_{h,h'} [\lambda_2 \delta_{c,c'} (1 - \delta_{p.p'}) + \lambda_3 \delta_{p,p'} (1 - \delta_{c,c'})] + \\ - 2\lambda_0 \delta_{h,h'} (1 - \delta_{c,c'}) (1 - \delta_{p,p'}) \\ h_{h,p,c} = -\lambda_1 (1 - 2N_{pc}) - \lambda_0 (1 - 2\overline{N})$$

2.3.3. Problem komiwojażera

Rozważmy N wierzchołkowy graf pełny[11] G z krawędziami ważonymi wagami $d_{i,j}$ oraz N dni.

- Kodowanie - bierzemy $N \cdot N$ jednostek (neuronów) kodujących drogę komiwojażera w następujący sposób

$$\sigma_{i,\nu} = \begin{cases} 1 & ; i\text{-te miasto zostało odwiedzone w } \nu \text{ dniu} \\ 0 & ; \text{w przeciwnym przypadku} \end{cases}$$

[10] W przypadku problemu planu zajęć już samo znalezienie stanu dopuszczalnego jest trudnym zadaniem, dlatego można rozważać sieć bez członu optymalizującego.

[11] Założenie pełności grafu wydaje się zbyt silne, jednak zawsze można je osłabić na odpowiednich krawędziach grafu pełnego wstawiając $d_{i,j} = \infty$, co symbolizuje nieobecność tej krawędzi w wyjściowym grafie.

Tylko do celów edukacyjnych

- Energia ograniczeń - w tym przypadku jest dość łatwa do sformułowania. Trzeba zadbać o to aby każde miasto było odwiedzone tylko raz, oraz aby w tym samym dniu komiwojażer nie odwiedzał dwóch miast.

$$E_{ogr} = \lambda_1 \left(\sum_i \left(\sum_\nu \sigma_{i,\nu} - 1 \right)^2 + \sum_\nu \left(\sum_i \sigma_{i,\nu} - 1 \right)^2 \right)$$

- Energia optymalizacji - musi być proporcjonalna do sumarycznej długości drogi (cyklu, gdyż jak wiadomo komiwojażer musi wrócić do punktu wyjścia).

$$E_{opt} = \lambda_2 \sum_{i,\nu} \sum_{j,\mu} d_{i,j} [\delta_{\nu,(\mu+1)} + \delta_{\nu,(\mu-1)}] \sigma_{i,\nu} \sigma_{j,\mu}$$

Co w sumie daje następujące wagi oraz pole zewnętrzne

$$w_{(i,\nu),(j,\mu)} = -\lambda_1 [\delta_{i,j}(1 - \delta_{\nu,\mu}) + \delta_{\nu,\mu}(1 - \delta_{i,j})] - \lambda_2 d_{i,j} (\delta_{\nu,(\mu+1)} + \delta_{\nu,(\mu-1)})$$
$$h_{i,\nu} = \lambda_1$$

2.4. Elementy maszyn Boltzmanna

Jak wspomniano w poprzednich rozdziałach, największą przeszkodą w osiąganiu przez sieci Hopfielda dobrych rozwiązań problemów jest podatność na popadania w minima lokalne. Jest to istotna okoliczność, gdyż okazuje się, że w skomplikowanych problemach sieć praktycznie zawsze kończy działanie w lepszym lub gorszym minimum lokalnym. Aby temu zaradzić trzeba wprowadzić mechanizm pozwalający wyjść z lokalnych basenów atrakcji, coś co pozwoli sieci przeskoczyć pomiędzy barierami potencjału, ale jej nie zdestabilizuje. Rozsądnym rozwiązaniem tych problemów okazała się być idea tak zwanych *maszyn Boltzmanna* zaproponowanych przez Hintona i Sejnowskiego w [17, 18, 2], polegająca na dodaniu do dynamiki sieci czynnika stochatycznego[12]. Aby uzyskać maszynę Boltzmanna wystarczy dodać do dynamiki sieci Hopfielda zasadę, iż w sytuacji, gdy spin neuronu zgadza się ze znakiem pola lokalnego, to z prawdopodobieństwem $e^{-2\beta|\Sigma_i|}$ zaburzamy stan tego neuronu nadając mu znak przeciwny[13]. W przypadku, gdy znak jest przeciwny zamieniamy odpowiednio stan neuronu jak w zwyczajnej dynamice. Później umotywujemy, że taki zabieg istotnie powoduje zadanie na przestrzeni stanów odpowiedniego rozkładu prawdopodobieństwa. Fakt, że wybrano rozkład taki a nie inny także będzie szerzej skomentowany w następnych rozdziałach, jednak warto tu wprowadzić kilka wstępnych pojęć[14]. Po pierwsze trzeba sobie zdać sprawę z tego, jakie konsekwencje niesie ze sobą dodanie czynnika losowego w ewolucji sieci. Tracimy mianowicie kilka użytecznych pojęć takich jak deterministyczna trajektoria w przestrzeni fazowej oraz atraktor (stan sieci zaburzany losowo nie utrzyma się długo w żadnym punkcie lub stabilnej pętli, dlatego pojęcie atraktora traci sens). Rolę punktów niezmienniczych ze względu na dynamikę przejmują rozkłady probabilistyczne (tak zwane rozkłady równowagowe, lub stacjonarne[15]). Znajomość rozkładu stacjonarnego a także rozpoznanie czy stany

[12] Zadającego na przestrzeni stanów rozkład probabilistyczny zwany w fizyce rozkładem Boltzmanna, stąd nazwa maszyn.

[13] Znaczenie czynnika β będzie opisane niżej, na tym etapie wystarczy myśleć o nim jako o pewnym parametrze.

[14] W sposób być może nieco mniej formalny, jednak formalizm fizyki statystycznej jest na tym etapie rozważań nieco zbyt rozbudowany, przez co może zaciemniać pewne aspekty rozważanych pojęć.

[15] W tym kontekście (układów termodynamicznych) wyróżnia się pewną szczególną klasę takich rozkładów, tak zwanych *miar Gibbsa*.

układu są rzeczywiście z takiego rozkładu losowane, staje się kluczowe w analizie badanych systemów. W przeciwieństwie do sieci deterministycznych w których mogło istnieć wiele aktraktorów dynamiki, w przypadku maszyn Boltzmanna dla zadanej sieci oraz średniej energii stanu[16], istnieje dokładnie jeden rozkład prawdopodobieństwa równowagowy, niezmienniczy na odpowiednią dynamikę Glaubera, który minimalizuje *entropię względną* względem rozkładu jednostajnego na przestrzeni fazowej. Rozkład taki nazywamy miarą Gibbsa. Wymaganie utrzymania średniej energii na zadanym poziomie skutkuje tym, że pojawia się dodatkowy parametr (β) zwany dalej temperaturą odwrotną (natomiast $T = \frac{1}{\beta}$ będziemy konsekwentnie nazywać temperaturą). Temperatura okazuje się być parametrem mierzącym w pewnym sensie losowość układu - w największym skrócie - gdy temperatura ($T = \frac{1}{\beta}$) jest wysoka (duża losowość), układ losuje stany których energia średnia (liczona po ewolucji, lub po realizacjach układu - *własność ergodyczności*) jest wysoka. Gdy temperatura jest niska (losowość mała) układ losuje stany których energia średnia jest odpowiednio niższa. Zależność pomiędzy temperaturą a energią nie jest jednak zupełnie prosta ze względu na pojawiający się czynnik entropii[17]. Najważniejszym elementem, który powoduje, że maszyny Boltzmanna są użyteczne jest pewna własność stacjonarnych miar Gibbsa (przy temperaturze zbiegającej do zera) - rozkłady te degenerują się w taki sposób, że całą swoją masę skupiają w punkcie będącym globalnym minimum funkcji energetycznej (rysunek 2.2). Ta okoliczność powoduje, że procedura (zwana *symulowanym wyżarzaniem*) polegającą na symulowanym przechodzeniu do zerowej temperatury (trzymając się rozkładu stacjonarnego) z prawdopodobieństwem jeden trafia w minimum globalne[18]. Dowód tego faktu jest zaskakująco prosty dlatego przytoczymy go razem z następującym twierdzeniem[19]:

Twierdzenie 2.4.0.1 *Niech $\mathbb{S}$ będzie skończoną przestrzenią stanów, $f : \mathbb{S} \to \mathbb{R}$ funkcją rzeczywistą ograniczoną o wartościach nieujemnych na tej przestrzeni (zatem posiadającą minimum i maximum). Niech $T > 0$ oznacza temperaturę natomiast $\alpha(T)$ oznacza prawdopodobieństwo tego, że element y wylosowany z rozkładu*

$$G(\overline{\sigma}) = \frac{e^{\frac{-f(\overline{\sigma})}{T}}}{\sum_{\overline{\sigma}' \in \mathbb{S}} e^{\frac{-f(\overline{\sigma}')}{T}}}$$

spełnia

$$y = \min_{\overline{\sigma} \in \mathbb{S}} f(\overline{\sigma})$$

Wtedy

$$\lim_{T \to 0} \alpha(T) = 1$$

Dowód 2.4.0.2 *Niech $\overline{\sigma}_l \in \mathbb{S}$; $l = 1...L$ będą minimalizatorami f oraz $f(\overline{\sigma}_l) = a$. Niech b będzie wartością drugiego w kolejności najlepszego minimum lokalnego. Element taki istnieje bo zbiór $\mathbb{S}$ jest skończony, chyba że funkcja f jest stała, wtedy jednak dowód staje się trywialny.*

$$b = \min_{\overline{\sigma}' \in \mathbb{S} \setminus \{\overline{\sigma}_l \ ; \ l=1...L\}} f(\overline{\sigma}')$$

[16] Zadajemy na rozkład ograniczenie, polegające na tym, że średnie energia sieci liczona względem niego ma mieć z góry zadaną wartość.

[17] Zostanie to wyjaśnione w dalszych rozdziałach.

[18] Tu trzeba zaznaczyć, że dynamika przy temperaturze zbiegającej do zera i dynamika w zerowej temperaturze to zupełnie różne rzeczy, dlatego nie da się zastąpić procedury przechodzenia z temperaturą do zera, przez wzięcie od razu układu w zerowej temperaturze.

[19] Jest to pewien przedsmak rozważań które będziemy prowadzić dalej.

Tylko do celów edukacyjnych

Ponieważ $a < b$, zatem

$$\lim_{T\to 0} e^{\frac{a-b}{T}} = 0$$

Niech $k = |\mathbb{S}|$ oraz $\alpha_l(T)$ będzie prawdopodobieństwem trafienia w l-ty minimalizator f ($l = 1 \dots L$). Wtedy

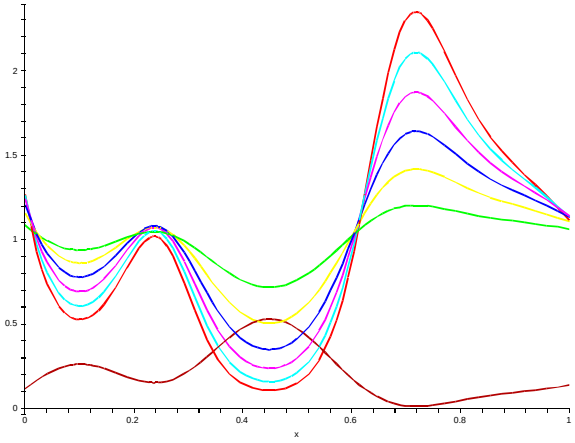
$$\alpha(T) = \sum_{l=1}^{L} \alpha_l(T)$$

oraz

$$\begin{aligned}\alpha_l(T) &= \frac{e^{\frac{-f(\overline{\sigma})}{T}}}{\sum_{\overline{\sigma}'\in\mathbb{S}} e^{\frac{-f(\overline{\sigma}')}{T}}} = \frac{e^{\frac{-a}{T}}}{\sum_{\overline{\sigma}'\in\mathbb{S}} e^{\frac{-f(\overline{\sigma}')}{T}}} = \\ &= \frac{e^{\frac{-a}{T}}}{L\cdot e^{\frac{-a}{T}} + \sum_{\overline{\sigma}'\in\mathbb{S}\setminus\{\overline{\sigma}_l\ ;\ l=1\dots L\}} e^{\frac{-f(\overline{\sigma}')}{T}}} \geq \frac{e^{\frac{-a}{T}}}{L\cdot e^{\frac{-a}{T}} + (k-L)e^{\frac{-b}{T}}} = \\ &= \frac{1}{L + (k-L)e^{\frac{a-b}{T}}} \to_{T\to 0} \frac{1}{L}\end{aligned}$$

Ponieważ minimalizatorów (globalnych) jest L zatem prawdopodobieństwo trafienia w dowolny z nich wynosi 1 (cała miara rozkłada się równomiernie na zbiorze minimalizatorów), ostatecznie $\alpha(T) \to_{T\to 0} 1$

□



Rysunek 2.2. Przykład funkcji (energetycznej) zadanej na odcinku $[0, 1]$ (czerwony wykres najniżej) oraz seria gęstości prawdopodobieństwa miar Gibbsa dla coraz większych parametrów β. Łatwo zaobserwować, że miary te coraz bardziej skupiają się na minimum funkcji energetycznej.

Źródło: Opracowanie własne

Dowód ten jest dość elegancki i nie wymaga zaawansowanych pojęć (dlatego został przytoczony w tym miejscu), jednak jego konsekwencje są znaczące, implikuje bowiem, że symulowane wyżarzanie efektywnie przeszukuje przestrzeń stanów i w granicy prowadzi do rozwiązania

Tylko do celów recenzyjnych

optymalnego[20]. Aby jednak głębiej przeanalizować właściwości stochastycznych układów spinowych (maszyn Boltzmanna) potrzebne są bardziej zaawansowane narzędzia matematyczne, łańcuchy Markowa oraz język fizyki statystycznej i termodynamiki.

Tylko do celów edukacyjnych

[20]Trzeba tu zaznaczyć, że symulowane wyżarzanie nie jest trywialną metodą optymalizacji globalnej i wymaga ostrożności. Nie można zbyt szybko przejść z temperaturą odwrotną β do ∞, trzeba bowiem mieć pewność osiągnięcia rozkładu stacjonarnego przy danej temperaturze (co w ogólności nie jest trywialne).

ROZDZIAŁ 3

Łańcuchy Markowa

3.1. Definicja i własności łańcuchów Markowa

Jak okaże się nieco dalej, maszyna Boltzmanna realizuje pewien specyficzny rodzaj procesu stochastycznego zwanego *łańcuchem Markowa* (czasem też procesem Markowa). Łańcuchy Markowa są istotnym komponentem znacznej większości opracowanych algorytmów stochatycznych. Aby wgłębić się nieco w opisywane pojęcia potrzeba jednak kilku definicji:

Definicja 3.1.0.1 *Procesem stochastycznym nazywamy ciąg zmiennych losowych* $(X_0, X_1, \ldots)$.

Jak widać definicja procesu stochastycznego jest niebywale prosta. Prostota ta jednak w rzeczywistości kryje dość skomplikowane i bogate pojęcie. Zdroworozsądkowo można myśleć o procesie stochatycznym jako o ciągu losowań, np. rzutach monetą lub kostką. Po krótkim zastanowieniu pojawiają się własności które proces może posiadać. Może np. się okazać, że zmienna X_n jest zależna (lub nie) od innych zmiennych w procesie, może się zdarzyć że rozkłady kolejnych zmiennych będą się zbliżać do pewnego rozkładu granicznego (jeśli ten rozkład graniczny będzie miał całą masę skupioną w jednym punkcie to sam ciąg $(X_0, X_1, \ldots)$ może też zbiec do deterministycznej granicy). Może się okazać, że od pewnego numeru rozkład zmiennych procesu się nie zmienia lub zaczyna oscylować itd. Jak widać jest mnóstwo możliwości które wymagają pewnego usystematyzowania. Ze względu na bogactwo opisywanych pojęć, zaczniemy od ograniczenia pola naszych rozważań do szczególnej sytuacji:

- Niech Ω będzie pewną przestrzenią probabilistyczną, zaś $\mathbb{S}$ pewnym skończonym zbiorem $|\mathbb{S}| = k$. Elementy $\mathbb{S}$ będziemy oznaczać $\overline{\sigma}_l$, $l = 1...k$.
- Będziemy rozważać rodzinę zmiennych losowych $X_l : \Omega \to \mathbb{S}$, gdzie l przebiega pewien przeliczalny zbiór indeksów (można przyjąć że $l \in \mathbb{N}$).

Definicja 3.1.0.2 *Proces stochastyczny będziemy nazywać łańcuchem Markowa jeśli dla każdego* $n \in \mathbb{N}$ *zachodzi własność*

$$\mathbb{P}(X_{n+1} = \overline{\sigma}_j | X_0 = \overline{\sigma}_{i_0}, \ldots, X_{n-1} = \overline{\sigma}_{i_{n-1}}, X_n = \overline{\sigma}_i) = \mathbb{P}(X_{n+1} = \overline{\sigma}_j | X_n = \overline{\sigma}_i)$$

Proces jest zatem procesem (łańcuchem) Markowa jeśli ma *własność zapominania*, to znaczy to jak będzie wyglądał następny element łańcucha zależy tylko i wyłącznie od jego bezpośredniego poprzednika (to jak będzie wyglądało "jutro" zależy wyłącznie od tego jak wygląda "dzisiaj", tak jakby całej przeszłości nie było[1]).

Definicja 3.1.0.3 *Łańcuch Markowa nazywamy jednorodnym jeśli dla dowolnej pary indeksów* i *oraz* j, $\mathbb{P}(X_{n+1} = \overline{\sigma}_j | X_n = \overline{\sigma}_i)$ *jest stałe i nie zależy od* n.

Oznacza to, że dla jednorodnych łańcuchów Markowa można utworzyć pojedynczą macierz zawierającą prawdopodobieństwa przejścia z jednego stanu do innego:

[1] Nie należy tego traktować jako ogólnej zasady rządzącej światem, a jedynie jako własność pewnego modelu matematycznego.

Tylko do celów edukacyjnych

Definicja 3.1.0.4 *Macierzą przejścia dla łańcucha Markowa* $(X_0, X_1, \ldots)$ *nazywamy macierz* $P_{i,j}$ *wymiaru* $k \times k$ *która spełnia*

$$P_{i,j} = \mathbb{P}(X_{n+1} = \overline{\sigma}_j | X_n = \overline{\sigma}_i)$$

Elementy $P_{i,j}$ *macierzy przejścia nazywamy prawdopodobieństwami przejścia ze stanu* $\overline{\sigma}_i$ *do* $\overline{\sigma}_j$*.*

Fakt 3.1.0.1 *Jeśli macierz* $P_{i,j}$ *jest macierzą przejścia pewnego łańcucha Markowa to spełnia następujące własności:*

- $$P_{i,j} \geq 0 \ ; \ i, j \in \{1, \ldots, k\}$$
- $$\sum_{j=1}^{k} P_{i,j} = 1 \ ; \ i \in \{1, \ldots, k\}$$

Użyteczne będzie także odnotowanie pewnego istotnego faktu:

Fakt 3.1.0.2 *Skoro istnieje dla danego procesu jednoznacznie wyznaczona macierz przejścia, to możemy skonstruować również graf* $G = (V, E)$*, którego zbiorem wierzchołków będzie zbiór stanów* $V = \mathbb{S}$*, zaś zbiorem krawędzi* $E = \{(\overline{\sigma}_i, \overline{\sigma}_j) \ ; \ P_{i,j} > 0\}$*. Krawędziom tego grafu możemy też od razu przyporządkować wagi* $w_{i,j} = P_{i,j}$*. Graf ten będziemy nazywać* grafem stanów *łańcucha Markowa.*

Definicja 3.1.0.5 *Stanem początkowym łańcucha (procesu) Markowa oznaczamy pewien wybrany stan* $\overline{\sigma} \in \mathbb{S}$ *od którego rozpoczynamy ewolucję (może on być wybrany losowo).*

Zbiór stanów w łańcuchu Markowa rozkłada się na dwie podstawowe klasy:

Definicja 3.1.0.6 *Stan* $\overline{\sigma}$ *nazywamy* przechodnim *jeśli po jednokrotnym trafieniu weń, można już nigdy do niego więcej nie wrócić, tzn.*

$$\mathbb{P}(X_s \text{ wraca do } \overline{\sigma} \text{ nieskończenie wiele razy} | X_t = \overline{\sigma}) = 0$$

Definicja 3.1.0.7 *Stan* $\overline{\sigma}$ *nazywamy* rekurencyjnym *jeśli prawdopodobieństwo powrotu do niego wynosi 1, tzn*

$$\mathbb{P}\left(\exists_{t'>t} X_{t'} = \overline{\sigma} | X_t = \overline{\sigma}\right) = 1$$

Ponadto, gdy stan jest rekurencyjny, może mieć pewną dodatkową własność:

Definicja 3.1.0.8 *Stan* $\overline{\sigma}$ *nazywamy* absorbującym *jeśli prawdopodobieństwo wyjścia z niego wynosi 0, innymi słowy*

$$\mathbb{P}(X_{n+1} = \overline{\sigma} | P_n = \overline{\sigma}) = 1$$

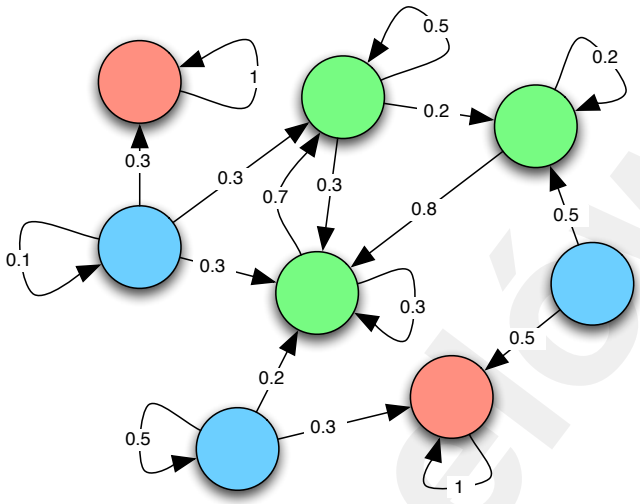
Warto tu odnotować, że każda grupa stanów rekurencyjnych jest atraktorem układu (atraktorem w sensie probabilistycznym, to znaczy po jednokrotnym wejściu do takiej grupy łańcuch z prawdopodobieństwem jeden się w niej utrzyma).

Definicja 3.1.0.9 Klasą rekurencyjną *nazywamy maksymalny zbiór stanów rekurencyjnych takich, że z każdego do każdego można przejść z niezerowym prawdopodobieństwem.*

Klasy rekurencyjne (jeśli istnieje ich więcej niż jedna) są zawsze rozłączne, co wynika z:

Fakt 3.1.0.3 *Jeśli dwie klasy rekurencyjne posiadają niezerowy przekrój to są równe.*

Tylko do celów edukacyjnych



Rysunek 3.1. Przykładowy łańcuch Markowa zawierający trzy stany przechodnie (niebieskie) dwa stany absorbujące (czerwone) i trzy stany tworzące klasę rekurencyjną (zielone). Etykiety na krawędziach oznaczają prawdopodobieństwa przejść ze stanu w którym krawędź się rozpoczyna do stanu gdzie krawędź się kończy.

Źródło: Opracowanie własne

3.2. Rozkłady stacjonarne

Rozważmy (podobnie jak w poprzednim rozdziale) łańcuch Markowa nad pewną skończoną przestrzenią stanów $\mathbb{S}$ z macierzą przejścia $P_{i,j}$.

Fakt 3.2.0.4 *Element $P^n_{i,j}$ określony*

$$P^n_{i,j} = \sum_{i_1,i_2,\ldots,i_{n-1}} P_{i,i_1} P_{i_1,i_2} \ldots P_{i_{n-1},j} = [P]^n_{i,j}$$

jest prawdopodobieństwem przejścia ze stanu $\overline{\sigma}_i$ do $\overline{\sigma}_j$ w n krokach. $[P]^n$ oznacza n-tą potęgę macierzy P.

Po mału ujawnia się wspomniana wcześniej elegancja i siła opisu łańcuchów Markowa, który sprowadza się do rachunku macierzowego (na pewnej podklasie zbioru macierzy kwadratowych - *macierzy stochatycznych*).

Definicja 3.2.0.10 *Rozkładem stacjonarnym łańcucha Markowa nazywamy miarę probabilistyczną $q = (q_1, \ldots, q_k)$ na przestrzeni stanów $\mathbb{S}$ taką, że jeśli zostanie zadana jako rozkład pierwszego elementu łańcucha X_0 powtórzy się jako rozkład drugiego elementu X_1.*

Rozkład stacjonarny jest zatem rozkładem niezmienniczym na ewolucję łańcucha. Biorąc pod uwagę powyższą definicję oraz podstawową własność łańcuchów Markowa, można dopowiedzieć, że jeśli rozkład q zostanie zadany na X_0 to powtórzy się na X_n dla dowolnego $n \in \mathbb{N}$. W dalszych rozważaniach spojrzymy na graf stanów łańcucha Markowa jak na sieć przepływową:

Tylko do celów edukacyjnych

Definicja 3.2.0.11 *Niech $t_{i,j} = q_i P_{i,j}$. Wyrażenie*

$$\sum_{i \neq j} t_{j,i}$$

będziemy nazywać dopływem do wierzchołka i, zaś

$$\sum_{i \neq j} t_{i,j}$$

odpływem z wierzchołka i.

Twierdzenie 3.2.0.3 *Rozkład q jest stacjonarny, gdy dla każdego i*

$$\sum_{i \neq j} t_{i,j} = \sum_{i \neq j} t_{j,i}$$

(dopływ i odpływ z każdego wierzchołka są równe)

Odnotujmy następujący

Fakt 3.2.0.5 *Po odwróceniu kolejności (kierunku biegu "czasu") w łańcuchu Markowa dostajemy także łańcuch Markowa. Niech*

$$\mathbb{P}(X_{n+1} = \overline{\sigma}_j | X_n = \overline{\sigma}_i) = P_{i,j}$$

oraz q będzie rozkładem stacjonarnym. Wtedy ze wzoru Bayesa

$$\mathbb{P}(X_n = \overline{\sigma}_j | X_{n+1} = \overline{\sigma}_i) = \frac{\mathbb{P}(X_n = \overline{\sigma}_j)}{\mathbb{P}(X_{n+1} = \overline{\sigma}_i)} \mathbb{P}(X_{n+1} = \overline{\sigma}_i | X_n = \overline{\sigma}_j) = \frac{q_j}{q_i} P_{j,i}$$

Definicja 3.2.0.12 *Łańcuch Markowa nazywamy odwracalnym jeśli istnieje rozkład początkowy q przy którym*

$$\mathbb{P}(X_{n+1} = \overline{\sigma}_j | X_n = \overline{\sigma}_i) = \mathbb{P}(X_n = \overline{\sigma}_j | X_{n+1} = \overline{\sigma}_i)$$

co jest równoważne z

$$q_i P_{i,j} = q_j P_{j,i}$$

Powyższe równanie nosi nazwę równania dokładnego bilansu.

Rozważając łańcuch Markowa jako sieć przepływową możemy wyobrazić sobie odwracalność jako sytuację w której po tej samej krawędzi dokładnie tyle samo prawdopodobieństwa wpływa do każdego wierzchołka co wypływa. Jest to silna własność, o czym świadczy twierdzenie:

Twierdzenie 3.2.0.4 *Jeśli rozkład q łańcucha $(X_0, X_1, \ldots)$ jest rozkładem dla którego spełnione są równania dokładnego bilansu, to jest też rozkładem stacjonarnym.*

Dowód 3.2.0.5 *Wystarczy sprawdzić, że dla każdego j*

$$q_j = \sum_{i=1}^{k} q_i P_{i,j}$$

Jednak

$$q_j = q_j \sum_{i=1}^{k} P_{j,i} = \sum_{i=1}^{k} q_j P_{j,i} = \sum_{i=1}^{k} q_i P_{i,j}$$

Tylko do celów edukacyjnych

□

Definicja 3.2.0.13 *Łańcuch Markowa nazywamy* nieprzywiedlnym *jeśli posiada jedną klasę rekurencyjną.*

Definicja 3.2.0.14 *Dla stanu $\overline{\sigma}_i$ określamy* okres stanu δ_i

$$\delta_i = \inf\{n > 0, (P^n)_{i,i} > 0\}$$

Definicja 3.2.0.15 *Okresem łańcucha nazywamy największy wspólny dzielnik okresów wszystkich stanów. Łańcuch nazwiemy* aperiodycznym *jeśli jego okres wynosi 1.*

Twierdzenie 3.2.0.6 *Niech łańcuch $(X_0, X_1, \ldots)$ będzie nieprzywiedlny oraz aperiodyczny. Wtedy istnieje dokładnie jeden stan stacjonarny i łańcuch do niego zbiega przy dowolnym stanie początkowym.*

Wyczerpujący dowód tego twierdzenia znajduje się w [14]. Nie jest przesadnie trudny, jednak dość długi.

Twierdzenie 3.2.0.7 *Niech $(X_0, X_1, \ldots)$ będzie łańcuchem Markowa ze stanem stacjonarnym q i macierzą przejścia P. Wtedy prawdziwe są następujące stwierdzenia:*

1. *q jest wektorem własnym P o wartości własnej 1.*
2. *Przy założeniu nieprzywiedlości i aperiodyczności q jest jedynym wektorem własnym o wartości własnej 1.*
3. *Promień spektralny macierzy P równa się* 1.

Dowód 3.2.0.8 *Prezentujemy tu jedynie zarysy dowodów:*

1. *$qP = q$ z definicji rozkładu stacjonarnego.*
2. *Poprzednie twierdzenie pokazuje, iż aperiodyczność i nieprzywiedlość implikuje istnienie dokładnie jednego rozkładu stacjonarnego q. Gdyby macierz P posiadała inny wektor własny q_1 dla wartości własnej 1, to wszystkie współrzędne w jego części rzeczywistej i wszystkie współrzędne w jego części urojonej musiałyby być tego samego znaku (zobacz argumentacja poniżej), zatem stosowna normalizacja q_1 byłaby rozkładem stacjonarnym dla rozważanego łańcucha. Tym samym q_1 musi być zespoloną wielokrotnością q, co należało wykazać.*
3. *Po pierwsze zauważmy, że macierz stochastyczna P ma wartości rzeczywiste. Oznacza to, że jeśli zadziałamy nią na wektor zespolony, to można osobno rozważać wyniki działania P na wartościach rzeczywistych i urojonych tego wektora. Niech $x = [x_1 + iy_1, x_2 + iy_2, \ldots, x_N + iy_N]$ będzie wektorem zespolonym. Niech $M(x)$ oznacza następujące wyrażenie*

$$M(x) = \sum_{j=1}^{N} |x_j + iy_j|$$

oraz

$$M^{\Re}(x) = \sum_{j=1}^{N} |x_j|$$

Tylko do celów edukacyjnych

i odpowiednio

$$M^{\Im}(x) = \sum_{j=1}^{N} |y_j|$$

Wtedy $M^{\Re}(xP) \leq M^{\Re}(x)$. *Mamy bowiem:*

$$M^{\Re}(xP) = \sum_{j=1}^{N} \left| \sum_{k=1}^{N} p_{k,j} x_j \right| \leq \sum_{j=1}^{N} \sum_{k=1}^{N} |p_{k,j} x_j| = \sum_{j=1}^{N} \sum_{k=1}^{N} p_{k,j} |x_j| = \sum_{j=1}^{N} |x_j| \sum_{k=1}^{N} p_{k,j} = \sum_{j=1}^{N} |x_j|$$

Analogicznie $M^{\Im}(xP) \leq M^{\Im}(x)$. *Załóżmy teraz, że istnieje taka zespolona wartość własna* λ *macierzy* P, *że* $|\lambda| > 1$. *Niech* x *będzie jej wektorem własnym. Wtedy oczywiście mamy:*

$$\lim_{n\to\infty} |xP^n| = \lim_{n\to\infty} |\lambda^n x| = \lim_{n\to\infty} |x_n| = \infty$$

z czego wynika, że

$$\lim_{x\to\infty} M(x_n) = \lim_{n\to\infty} \sum_{j=1}^{N} |x_j^{(n)} + i y_j^{(n)}| = \infty$$

To z kolei implikuje, że

$$\lim_{n\to\infty} M^{\Re}(x_n) = \infty$$

lub

$$\lim_{n\to\infty} M^{\Im}(x_n) = \infty$$

Co stoi w sprzeczności z faktem $M^{\Re}(xP) \leq M^{\Re}(x)$ *i* $M^{\Im}(xP) \leq M^{\Im}(x)$. *Nie może być zatem wartości własnej* λ *takiej, że* $|\lambda| > 1$

□

Definicja 3.2.0.16 *Niech* $\eta(x) = x\ln(x)$, *przy czym przyjmujemy* $\eta(0) = 0$. *Niech* π, θ *będą rozkładami probabilistycznymi na* $\mathbb{S}$. *Określamy*

$$\mathbb{H}(\theta|\pi) = \sum_{i=1}^{k} \theta_i \ln\left(\frac{\theta_i}{\pi_i}\right) = \sum_{i=1}^{k} \eta\left(\frac{\theta_i}{\pi_i}\right) \pi_i$$

$\mathbb{H}$ będziemy nazywać entropią względną rozkładu θ względem π.

Twierdzenie 3.2.0.9 *Niech* q *będzie stanem stacjonarnym łańcucha o macierzy przejścia* P. *Wtedy*

$$\mathbb{H}(\theta P|q) \leq \mathbb{H}(\theta|q)$$

(o ile obie strony równania mają sens).

Tylko do celów edukacyjnych

Dowód 3.2.0.10

$$\begin{aligned}
\mathbb{H}(\theta P|q) &= \sum_{i=1}^{k} q_i \eta\left(\frac{1}{q_i}\sum_{j=1}^{k}\theta_j P_{i,j}\right) = \\
&= \sum_{i=1}^{k} q_i \eta\left(\sum_{j=1}^{k}\frac{\theta_j}{q_j}\frac{q_j P_{i,j}}{q_i}\right) \leq \quad \textit{(Tu korzystamy z nierówności Jensena)} \\
&\leq \sum_{i=1}^{k} q_i \sum_{j=1}^{k}\frac{q_j P_{i,j}}{q_i}\eta\left(\frac{\theta_j}{q_j}\right) = \\
&= \sum_{i=1}^{k}\sum_{j=1}^{k} q_j P_{j,i}\eta\left(\frac{\theta_j}{q_j}\right) = \\
&= \sum_{j=1}^{k} q_j \eta\left(\frac{\theta_j}{q_j}\right) = \mathbb{H}(\theta|q)
\end{aligned}$$

Pokażemy teraz, że maszyna Boltzmanna (wstępnie opisana w 2.4) jest łańcuchem Markowa, który posiada stan stacjonarny. Przypomnijmy krótko rozważany model:

- Mamy sieć Hopfielda o N neuronach przyjmujących stany 1 lub -1 oraz parametr β zwany temperaturą odwrotną
- Przestrzeń stanów $\mathbb{S} = 2^N$
- Przestrzegamy konwencji notacyjnej: σ_i oznacza stan neuronu o numerze i, zaś $\overline{\sigma}_i$ oznacza element przestrzeni $\mathbb{S}$ indeksowany przez i. [2]
- Dynamika sieci wygląda następująco
  - Losujemy jednostkę σ_i
  - Jeśli spin σ_i jest niezgodny z lokalnym polem wypadkowym Σ_i ustawiamy go zgodnie $\sigma_i = sgn(\Sigma_i)$.
  - Jeśli spin σ_i jest zgodny z lokalnym polem wypadkowym to z prawdopodobieństwem $e^{-2\beta|\Sigma_i|}$ ustawiamy mu znak przeciwny.

Twierdzenie 3.2.0.11 *Dla układu zdefiniowanego powyżej rozkład stacjonarny ma postać*

$$\mathbb{P}(\overline{\sigma}) = \frac{e^{-\beta E(\overline{\sigma})}}{\sum_{\overline{\sigma}'} e^{-\beta E(\overline{\sigma}')}} = \frac{e^{-\beta E(\overline{\sigma})}}{Z[\beta]} \tag{3.1}$$

Dowód 3.2.0.12 *Pokażemy nieco więcej, mianowicie to, że układ spełnia równanie dokładnego bilansu, a zatem łańcuch jest odwracalny. Rozważmy konfiguracje $\overline{\sigma}$ i $\overline{\sigma}'$ różniące się jedynie na i-tym miejscu σ_i. Niech $\overline{\sigma}$ będzie zgodne w σ_i z polem Σ_i zaś $\overline{\sigma}'$ nie. Wtedy przepływ prawdopodobieństwa z $\overline{\sigma}$ do $\overline{\sigma}'$ wynosi*

$$\mathbb{P}(\overline{\sigma}')\frac{1}{N}$$

[2] Trzeba tu uważać - we wszystkich rozważaniach dotyczących łańcuchów Markowa oznaczaliśmy elementy $\mathbb{S}$ przez $\overline{\sigma}$. Konsekwentnie tak będziemy oznaczać stany układu, których nie należy mylić ze stanem pojedynczych neuronów.

Tylko do celów edukacyjnych

zaś przepływ z $\overline{\sigma}$ do $\overline{\sigma}'$

$$\mathbb{P}(\overline{\sigma})\frac{1}{N}e^{-2\beta|\Sigma_i|}$$

Sprawdzamy czy

$$\begin{aligned}
\mathbb{P}(\overline{\sigma}')\frac{1}{N} =& \mathbb{P}(\overline{\sigma})\frac{1}{N}e^{-2\beta|\Sigma_i|} \\
\Updownarrow & \\
\frac{e^{-\beta E(\overline{\sigma}')}}{Z[\beta]} =& \frac{e^{-\beta E(\overline{\sigma})}}{Z[\beta]}e^{-2\beta|\Sigma_i|} \\
\Updownarrow & \\
e^{-\beta E(\overline{\sigma}')} =& e^{-\beta(E(\overline{\sigma}+2|\Sigma_i|))} \\
\Updownarrow & \\
E(\overline{\sigma}) =& E(\overline{\sigma}) + 2|\Sigma_i|
\end{aligned}$$

O ostatniej równości już wiemy, że jest prawdziwa (porównaj dowód 2.1.0.2). Zatem dla rozkładu (3.1) spełnione są równania dokładnego bilansu, zatem rozkład ten jest stacjonarny. Co więcej jest to jedyny rozkład stacjonarny, gdyż łańcuch jest nieprzywiedlny [3]*, a ponieważ jest także aperiodyczny*[4] *to zawsze do niego zbiega.*

Jeśli teraz porównamy uzyskany wynik z twierdzeniem 2.4.0.1, uzyskujemy motywację dlaczego w rozdziale 2.4 wybrano taki a nie inny mechanizm stochastyczny.

3.3. Algorytm doskonały Proppa-Wilsona

Istotnym problemem symulacji łańcuchów Markowa jest określenie czasu po którym zbiega on do rozkładu stacjonarnego. W dużej ilości zastosowań symuluje się łańcuch przez dostatecznie długi okres czasu (odpowiednią ilość przejść stanów) w nadziei, że wystarczy to do osiągnięcia rozkładu bliskiego stacjonarnemu. W ogólności jednak może to trwać bardzo długo z powodu olbrzymiej przestrzeni konfiguracji. Proste symulowanie niestety nie przekazuje nam żadnej informacji kiedy nastąpiło zbliżenie do rozkładu stacjonarnego, nie daje to warunku stopu ani gwarancji poprawności uzyskanej próbki. Rozwiązanie tego problemu zaproponowali J. Propp i D. Wilson w [32]. Algorytm ich autorstwa w swojej pierwotnej wersji prezentuje dość ciekawe podejście polegające na tym, że nie symulujemy ewolucji łańcucha w przyszłość, zamiast tego zaczynamy od pewnego punktu w przeszłości i symulujemy łańcuch do punktu $t = 0$, badając przy okazji czy zachodzi w nim pewna własność (o której za chwile). Zanim jednak zaczniemy dokładniej opisywać algorytm Proppa-Wilsona, musimy zacząć od pewnych wstępnych definicji:

Definicja 3.3.0.17 *Niech $\overline{\sigma}_i$ będzie stanem łańcucha Markowa, z którego do innych można przejść prawdopodobieństwami $P_{i,1}, P_{i,2}, \ldots, P_{i,k}$. Niech U będzie zmienną losową z rozkładu jednostajnego na odcinku $[0,1]$. Funkcją przejścia dla łańcucha nazywamy taką funkcję $\varphi : \mathbb{S} \times [0,1] \to \mathbb{S}$, która spełnia następującą zależność:*

$$\forall_{i,j}\ \mathbb{P}(\varphi(\overline{\sigma}_i, U) = \overline{\sigma}_j) = P_{i,j}$$

[3] Łatwo to sprawdzić, wystarczy pokazać, że z każdego do każdego innego stanu można przejść z niezerowym prawdopodobieństwem.

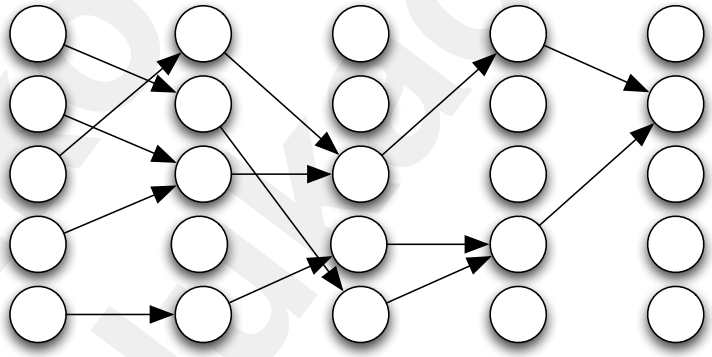
[4] To widać od razu, bo okres większości stanów wynosi oczywiście 1 (wystarczy wskazać jeden taki stan).

Funkcja przejścia jest zatem deterministyczną funkcją która przekształca miarę jednostajną na odcinku na rozkład prawdopodobieństw warunkowych w łancuchu Markowa. Pozornie może się wydawać, że taką funkcję trudno skonstruować, zazwyczaj jednak okazuje się to dość proste. Zmienna U jest tutaj jedynym źródłem losowości. W rozważanym wcześniej przykładzie maszyny Boltzmanna z U można uzyskać numer neuronu do zmiany, a następnie umiejętnie odwracając rozkład wykładniczy dowiedzieć się czy powinniśmy wykonywać zamianę spinu wbrew polu lokalnego. W praktyce programistycznej stosuje się w tym miejscu zazwyczaj dwa losowania jednak można pokazać, że zawsze można sprawę załatwić jednym. W tej sytuacji ewolucja łańcucha Markowa jest zdeterminowana za pośrednictwem funkcji przejścia φ poprzez ciąg zmiennych losowych $U_1, U_2, \ldots$ z rozkładu jednostajnego na $[0, 1]$. W tym momencie jesteśmy prawie gotowi do zaprezentowania algorytmu Proppa-Wilsona. Niech $N_1, N_2, N_3, \ldots$ będzie ściśle rosnącym ciągiem liczb naturalnych. Niech $u_1, u_2, u_3, \ldots$ będzie ustaloną sekwencją liczb wylosowanych niezależnie z rozkładu jednostajnego na odcinku $[0, 1]$.*Algorytm Proppa-Wilsona:*

1. Ustal wartość $i := 1$

2. Symuluj ewolucję łańcucha od momentu $t = -N_i$ równolegle dla każdego stanu $\overline{\sigma} \in \mathbb{S}$ wybranego jako stan początkowy, używając sekwencji liczb $u_{N_i}, u_{N_i-1}, \ldots, u_1$ w mechanizmie przejścia. Jeśli podczas pewnej iteracji dwie instancje łańcucha trafią do tego samego stanu, sklej je, symulując dalej tylko jeden z nich.

3. Jeśli w momencie $t = 0$ pozostała tylko jedna instancja łańcucha (reszta się skleiła) to zakończ pracę. Uzyskany stan końcowy jest wylosowany ze stanu stacjonarnego. Jeśli pozostało więcej łańcuchów ustal $i = i + 1$ i przejdź do punktu 2.



Rysunek 3.2. Schematyczny przebieg algorytmu Proppa-Wilsona na przestrzeni fazowej o 5 stanach. W kolejnych krokach równolegle liczone drogi ewolucji się sklejają, by ostatecznie zakończyć się w jednym stanie. Stan ten zgodnie z istotą algorytmu jest wylosowany z rozkładu stacjonarnego.

Źródło: Opracowanie własne

Algorytm jest bardzo prosty w sformułowaniu, trzeba jednak nadmienić tutaj o kilka sprawach:

- Na początku symulacji mamy startować z każdego $\overline{\sigma} \in \mathbb{S}$. Warunek ten oczywiście w większości przypadków nie może być spełniony ze względu na rozmiar przestrzeni

konfiguracji $\mathbb{S}$. W praktyce można stosować pewną próbkę stanów jednostajne wybraną po przestrzeni $\mathbb{S}$, algorytm jednak w tym momencie staje się algorytmem przybliżonym.

- Sekwencja $u_1, u_2, \ldots$ musi być pamiętana w czasie wykonywania algorytmu. Może być to kłopotliwe, gdy ilość pamiętanych liczb zacznie przekraczać dostępną pamięć.[5] W praktyce często pamięta się tylko ziarna (*zalążki - seeds*) generatora pseudolosowego, który po zainicjowaniu odpowiednim ziarnem generuje ustaloną sekwencję liczb. Praktyka ta jednak z wielu powodów nie jest polecana, ze względu na niedoskonałości generatorów pseudolosowych[6].

- Algorytm P-W jest bardzo przyjemny do urównoleglenia - każda maszyna komputera równoległego może niezależnie symulować jeden z łańcuchów. Komunikacja między procesorami nie jest duża - sprowadza się do rozpropagowania sekwencji liczb losowych (w przypadku każdego łańcucha musi być użyta ta sama sekwencja) oraz uzgodnienia raz na pewien czas które z symulowanych łańcuchów trzeba skleić. Mając zatem komputer równoległy z N procesorami bardzo łatwo można zaimplementować przybliżony algorytm P-W startujący z N losowo wybranych stanów początkowych.

W pewnych szczególnych przypadkach algorytm $P-W$ może mieć bardzo elegancką formę: jeśli na przestrzeni stanów $\mathbb{S}$ daje się zdefiniować odpowiedni porządek częściowy, wystarczy symulować tylko dwa łańcuchy zaczynające się w skrajnych punktach $\mathbb{S}$ względem tego porządku. Podamy tutaj definicję w dość szczegółowy sposób odnoszącą się do układów spinowych (w ogólności zagadnienie jest nieco bardziej skomplikowane)

Brak precyzji poprzedniego zdania rozwieją następujące definicje:

Definicja 3.3.0.18 *Rozważmy N elementowy układ spinowy w którym pojedyncze elementy przyjmują wartości ze zbioru $\{-1, 1\}$ oraz pewną dynamikę. Powiemy, że dynamika ta jest monotoniczna jeśli dla każdych dwóch konfiguracji $\overline{\sigma}$ i $\overline{\eta}$ zachodzi:*

$$\forall_{\overline{\sigma} \geq \overline{\eta}} \quad \mathbb{P}(\sigma_i' = 1 | \overline{\sigma}) \geq \mathbb{P}(\eta_i' = 1 | \overline{\eta})$$

Jeśli pokażemy, że dynamika stochastyczna zachowuje porządek w sensie jak powyżej (po współrzędnych), to wystarczy w algorytmie $P-W$ wybrać skrajne stany początkowe $[-1, -1, \ldots, -1]$ oraz $[1, 1, \ldots, 1]$ i symulować do czasu sklejenia jedynie dwa łańcuchy. Niestety przypadki dynamik monotonicznych w ogólności nie są częste.

[5] W pracy [37] D. Wilson opisał pewną modyfikację algorytmu P-W, w której stosuje się każdą liczbę losową tylko raz.

[6] Wymaga to nieco szerszego komentarza - generatory losowe popadają w cykle, które w codziennych zastosowaniach nie są zauważalne, jednak w symulacji algorytmem P-W (szczególnie w długich sekwencjach) mogą się ujawnić. Pewnym wyjściem jest zastosowanie mechanizmów generowania liczb losowych przez jądra systemów Uniksowych (urządzenie */dev/random*) które z założenia ma generować ciągi w sposób niedeterministyczny, pobierając jako źródło losowości szumy występujące w sprzęcie komputerowym. Można także korzystać z fizycznych generatorów losowości, lub gotowych danych losowych dostępnych w Internecie.

Tylko do edukacyjnych

ROZDZIAŁ 4

Aparat termodynamiki

W tym rozdziale zajmiemy się pewnymi elementami aparatu matematycznego termodynamiki, czyli fizyki statystycznej. Na początek należy wspomnieć o motywacjach kryjących się za chęcią opisywania układów fizycznych za pomocą statystyki. Wyobraźmy sobie mianowicie hipotetyczny układ kulek które latają w pewnej ograniczonej przestrzeni, odbijających się od siebie i od ścian w sposób doskonale sprzężysty. Niech pomiędzy tymi kulkami (i otoczeniem) działa pewne oddziaływanie (np. grawitacja). Ruch kulek w takim układzie opisywany jest dokładnie poprzez równania ruchu Newtona, teoretycznie więc wszelkie cechy ewolucji są znane. Okazuje się jednak, że jeśli kulek będzie dużo (zasadniczo więcej niż dwie), to rozwiązanie równań ruchu staje się nietrywialne[1] (słynny problem trzech ciał). Co więcej, gdy kulek będzie bardzo dużo, układ zacznie wykazywać pewne *cechy kolektywne*, które nie wynikają w sposób prosty z równań ruchu (wydaje się, że wynikają z samego faktu bycia układem kolektywnym). Badaniem takich *cech kolektywnych* zajmuje się właśnie fizyka statystyczna. Należy tu wspomnieć o pewnych zasadniczych różnicach w stosunku do opisu za pomocą mechaniki klasycznej:

- Nie interesuje nas los pojedynczej cząsteczki (elementu).
- Nie interesują nas dokładne wartości pewnych parametrów (funkcji) układu, a jedynie ich wartości średnie przy zadanych ograniczeniach.
- Jeśli interesują nas wartości średnie parametrów, to interesują nas także rozkłady prawdopodobieństwa na przestrzeni fazowej układu (względem których te średnie powstają).
- Zakładamy, że układ, który ma bardzo dużo stopni swobody jest dobrze przybliżany przez model nieskończony.[2]

Co ciekawe, często okazuje się, że różne układy wykazują w pewien sposób analogiczne cechy kolektywne mimo, że na poziomie molekularnym rządzą nimi zupełnie inne zasady. Można zatem wnioskować, że samo bycie układem składającym się z wielu odziałujących ze sobą cząstek skutkuje pewnymi charakterystycznymi zachowaniami, wynikającymi z jakichś fundamentalnych zasad statystyki, którym przyroda jest wierna. Jest to dodatkowa motywacja, by badać układy kolektywne, które niekoniecznie mają odpowiedniki w świecie fizycznym. Pewne cechy takich hipotetycznych układów mogą przenosić się na układy rzeczywiste (takim przykładem są na przykład przejścia fazowe opisane dalej w tym rozdziale). Język opisu układów kolektywnych (w założeniu elementów dość prostych - cząsteczek, atomów, spinów magnetycznych) znajduje także zastosowanie przy opisie systemów, w których podstawowy element może być czymś znacznie bardziej skomplikowanym a oddziaływania znacznie bardziej wyrafinowane. Przykładami takich elementów podstawowych może być zbiór neuronów

[1] Niemożliwe analitycznie, a i numerycznie trudne.

[2] To stwierdzenie to tak naprawdę esencja fizyki statystycznej. Większość modeli termodynamicznych formułuje się w przypadkach skończonych a następnie przechodzi do tak zwanej granicy termodynamicznej (do nieskończoności z rozmiarem układu). Wtedy pewne rzeczy opisuje się łatwiej niż w przypadku układów skończonych.

Tylko do celów edukacyjnych!

(np. sieć Hopfielda - takie zastosowanie opisane jest w tej pracy), czy nawet złożony organizm żywy (opis na przykład roju pszczół, tłumu ludzi etc.). Zaczniemy jednak od pewnych zupełnie klasycznych układów, na których wyrosła dzisiejsza fizyka statystyczna.

4.1. Przykłady układów opisywanych za pomocą fizyki statystycznej

4.1.1. Klasyczny gaz

Niech $\Lambda \subset \mathbb{R}^3$ będzie ograniczonym podzbiorem trójwymiarowej przestrzeni rzeczywistej o granicach będących powierzchniami prawie wszędzie różniczkowalnymi. Rozważmy N identycznych punktów (cząstek) poruszających się po Λ. Każda cząstka jest charakteryzowana przez jej pozycję $q_i \in \Lambda$ oraz prędkość $v_i\mathbb{R}^3$. Stan układu jest opisywany poprzez sekwencję

$$((q_1, v_1), (q_2, v_2), \ldots, (q_N, v_N)) \equiv (Q, V)$$

Jednak z pewnych przyczyn wygodniej jest używać pędu $p = mv$ niż prędkości:

$$((q_1, p_1), (q_2, p_2), \ldots, (q_N, p_N)) \equiv (Q, P)$$

Przestrzeń fazowa takiego układu to

$$\Omega_{\Lambda,N} = (\Lambda \times \mathbb{R}^3)^N \subset (\mathbb{R}^3 \times \mathbb{R}^3)^N = \mathbb{R}^{6N}$$

Energię (zwaną czasem Hamiltonianem) stanu (Q, P) wyraża się za pomocą wzoru

$$H_{\Lambda,N}(Q, V) = \sum_{i=1}^{N} \frac{mv_i^2}{2} + \sum_{1 \le i < j \le N} U(q_i - q_j) + \sum_{i=1}^{N} V_b(q_i)$$

Pierwszy człon Hamiltonianu odpowiada za energię kinetyczną, drugi za energię potencjalną oddziaływań pomiędzy cząsteczkami, trzeci za energię tak zwanego potencjału brzegowego, określającego sposób oddziaływania cząsteczek z brzegami dziedziny Λ. Potencjał U spełnia w modelu gazu klasycznego kilka rygorystycznych ograniczeń:

- Zakładamy że potencjał U ma lokalny charakter, zatem jest funkcją szybko malejącą (możemy założyć że ma nośnik ograniczony).
- Przez R będziemy oznaczań promień oddziaływań.
- Potencjał U często jest inwariantny na obroty, zatem $U(q_i - q_j) = U(|q_i - q_j|)$.

W praktyce fizycznej często korzysta się z potencjału Lennarda-Jonesa danego wzorem:

$$U(r) = \frac{A}{r^{12}} - \frac{B}{r^6}$$

dla pewnych parametrów A i B (zależnych od modelu), przyciętego tak aby powyżej pewnego R przyjmował wartość 0. Dynamikę układu opisują równania ruchu Newtona:

$$\frac{dq_i}{dt} = v_i$$

$$m\frac{dv_i}{dt} = \sum_{i \neq j} F(q_j - q_i) + F_b(q_i), \quad i = 1, 2, \ldots, N$$

Tylko do celów edukacyjnych

gdzie

$$F(q_j - q_i) = -(\nabla U)(q_j - q_i), \quad F_b(q_i) = -(\nabla V_b)(q_i)$$
$$-\nabla U = \left\{ -\frac{\partial U}{\partial q^{(1)}}, -\frac{\partial U}{\partial q^{(2)}}, -\frac{\partial U}{\partial q^{(3)}} \right\}$$

Równania ruchu wraz z zasadami odbicia od brzegów domeny generują ruch w przestrzeni fazowej (rodzinę przekształceń $\Omega_{\Lambda,N} \to \Omega_{\Lambda,N}$). Każdy punkt $(Q_0, P_0) \in \Omega_{\Lambda,N}$ po czasie t przechodzi po krzywej całkowej równań ruchu do punktu (Q_t, P_t). Rodzinę takich przekształceń przestrzeni fazowej w siebie będziemy nazywać dynamiką i oznaczać S_t. Dynamika ma następujące własności:

- Zachowuje 6-N wymiarową objętość (*zasada zachowania objętości w przestrzeni fazowej*)
$$Vol_{6N}(S_t A) = Vol_{6N}(A), \quad A \subset \Omega_{\Lambda,N}$$
Właśnie tutaj ujawnia się wygoda zastosowania w opisie stanu pędu a nie samej prędkości. Zasada ta nie była by prawdziwa dla przestrzeni fazowej biorącej pod uwagę jedynie prędkość, zjawiska związane z różnicami mas poszczególnych cząstek powodowałyby rozmaite deformacje objętości. Można jednak założyć, że masy cząstek są wszystkie równe jedności. W takim uproszczonym modelu zasada zachowania objętości odnosi się także do przestrzeni fazowej położenia i prędkości.

- Zachowuje energię (*zasada zachowania energii*)
$$H_{\Lambda,N}(S_t(Q_0, V_0)) = H_{\Lambda,N}((Q_0, V_0))$$

Jest zatem zasadne rozważanie powierzchni o stałej energii

$$\Omega_{\Lambda,N,E} = \{(Q, V) \in \Omega_{\Lambda,N} : H_{\Lambda,N}(Q, V) = E\}$$

które są niezmiennicze ze względu na dynamikę

$$S_t \Omega_{\Lambda,N,E} = \Omega_{\Lambda,N,E}$$

4.1.2. Gaz konfiguracyjny - model matematyczny

Gaz konfiguracyjny to pewne uproszczenie modelu gazu klasycznego, w którym zaniedbujemy obecność składowej prędkości. Interesują nas tylko pozycje cząsteczek oraz ich wzajemne energie potencjalne. Takie uproszczenie wynika z odejścia od fizyki prawdziwych układów, a skupieniu się jedynie na pewnych matematycznych własnościach. Stan takiego układu opisywany jest ciągiem położeń:

$$Q = (q_1, q_2, \ldots, q_N), \quad q_i \in \Lambda, \quad i = 1...N, \quad \Lambda \subset \mathbb{R}^3$$

zaś przestrzeń fazowa

$$\Lambda^N = \Lambda \times \Lambda \times \ldots \times \Lambda \equiv \Omega^{conf}_{\Lambda,N} \subset \mathbb{R}^{3N}$$

Energię wyliczamy analogicznie jak w przypadku modelu gazu klasycznego, omijając czynnik odpowiedzialny za energię kinetyczną:

$$H_{\Lambda,N}(Q) = \sum_{1 \le i < j \le N} U(q_i - q_j) + \sum_{i=1}^{N} V_b(q_i)$$

W takim modelu nie ma dynamiki, nie przeszkadza to jednak w rozważaniu powierzchni o stałej energii

$$\Omega^{conf}_{\Lambda,N,E} = \{(Q) \in \Omega^{conf}_{\Lambda,N} : H_{\Lambda,N}(Q) = E\}$$

Tylko do celów edukacyjnych

4.1.3. Zbiór spinów magnetycznych, model Isinga

Niech $\Lambda \subset \mathbb{Z}^d$ będzie skończonym podzbiorem d wymiarowej kraty całkowitoliczbowej. Rozważmy zbiór Ω_Λ funkcji postaci

$$\sigma = \{\sigma(x), x \in \Lambda\}$$

które przyjmują wartości[3] w zbiorze $\{-1, 1\}$. Przestrzeń Ω_Λ jest przestrzenią fazową systemu, który fizycznie opisuje takie zjawiska jak na przykład magnetyzm. Każdy punkt zbioru Λ odpowiada elementarnemu spinowi magnetycznemu, zaś jego wartościowanie funkcją σ odpowiada zwrotowi tego spinu. Podobnie jak w przypadku gazu możemy zadać potencjał U, który określi jak poszczególne spiny na siebie wpływają. Energia konfiguracji $\sigma = \{\sigma(x), x \in \Lambda\}$ ma formę

$$H_\Lambda(\sigma) = \sum_{x,x' \in \Lambda, x \neq, x'} U(x - x')\sigma(x)\sigma(x') - h \sum_{x \in \Lambda} \sigma_x \tag{4.1}$$

W najprostszym przypadku, potencjał $U(s)$ znika dla $|s| > 1$, zatem spiny oddziaływują tylko ze swoimi najbliższymi sąsiadami. Model taki nazywamy modelem Isinga. Co ciekawe model ten został zaproponowany przez Wilhelma Lenza około roku 1920 w nadziei zrozumienia zjawiska ferromagnetyzmu. Było to już po odkryciach P. Curie które wskazywały, iż zjawisko ferromagnetyzmu znika w wysokiej temperaturze (temperatura Curie). Zjawiska te nie były rozumiane, Lenz zadał opisany wyżej model do analizy dla swojego doktoranta Ernsta Isinga. Ising rozważył model jednowymiarowy i pokazał, że nie występują w nim przejścia fazowe, z czego (błędnie) wywnioskował, że model ten nie nadaje się do opisu ferromagnetyzmu[4]. Później okazało się, że przejścia fazowe występują w modelu Isinga dla wymiarów wyższych niż 1...

4.2. Rozkłady mikrokanoniczne i kanoniczne (miary Gibbsa)

Jak już było wspomniane wcześniej, idea fizyki statystycznej polega na wnioskowaniu o wartościach średnich parametrów pewnych skomplikowanych układów fizycznych. Wartości średnie muszą być liczone względem pewnych rozkładów prawdopodobieństwa, stąd centralnym zagadnieniem tej teorii jest konstrukcja i badanie własności rozkładów probabilistycznych na przestrzeniach fazowych. Szczególne znaczenie mają w tym kontekście rozkłady inwariantne na dynamikę, gdyż w takich przypadkach może zajść własność ergodyczności układu[5] (średnia liczona względem takiego rozkładu po przestrzeni fazowej równa się średniej liczonej po czasie na trajektorii dynamiki). Rozważania zaczniemy od dość elementarnej konstrukcji konstrukcji miary, dziedziczonej po mierze Lebesgue'a (objętości) na przestrzeni fazowej, zwanej czasem miarą Gelfanda-Leraya. Rozważmy powierzchnię o stałej energii $\Omega_{\Lambda,N,E}$. Definiujemy miarę na podzbiorach $B \subset \Omega_{\Lambda,N,E}$ w następujący sposób

$$\nu_{\Lambda,N,E}(B) = \lim_{\Delta E \to 0} \frac{Vol(\Delta B)}{\Delta E} \tag{4.2}$$

ΔB to w tym przypadku wycinek przestrzeni fazowej mieszczący się pomiędzy powierzchniami $\Omega_{\Lambda,N,E}$ a $\Omega_{\Lambda,N,E+\Delta E}$, wychodzący prostopadle z powierzchni $\Omega_{\Lambda,N,E}$. Miara ta istnieje ze względu na następujące

[3] W pełnej ogólności mogą przyjmować wartości w dowolnej przestrzeni topologicznej, zbiór $\{-1, 1\}$ jest jednak bardzo wygodny do analizy.

[4] Jest to pewna nauczka aby pochopnie nie wyciągać daleko idących wniosków. W efekcie model Isinga odniósł potężny sukces naukowy, dobrze opisując zjawisko ferromagnetyzmu.

[5] Choć zdecydowanie nie musi. Generalnie jeśli o jakimś układzie da się powiedzieć, że jest ergodyczny, to jest to powód do radości, ze względu na arsenał twierdzeń które w takich (i tylko takich) układach są prawdziwe.

Twierdzenie 4.2.0.1 *Niech $f(x), x \in \mathbb{R}^n$ będzie gładką funkcją o skończonej ilości punktów krytycznych (takich w których $f'(x) = 0$). Dla dowolnej powierzchni spełniającej równanie $f(x) = E$ i dowolnego podzbioru tej powierzchni granica 4.2 istnieje.*

Dowód tego twierdzenia można znaleźć w literaturze [12]. Jeśli teraz znormalizujemy[6] miarę $\nu_{\Lambda,N,E}$ uzyskami *rozkład mikrokanoniczny* na powierzchni energetycznej $\Omega_{\Lambda,N,E}$:

$$P^{micro}_{\Lambda,N,E}(A) = \frac{\nu_{\Lambda,N,E}(A)}{\nu_{\Lambda,N,E}(\Omega_{\Lambda,N,E})}$$

Jak pamiętamy w modelu gazu klasycznego dynamika zachowywała energię oraz objętość w przestrzeni fazowej, zatem także miary Gelfanda-Leraya i rozkład mikrokanoniczny są inwariantne ze względu na tę dynamikę. Miara mikrokanoniczna posiada jednak szereg wad, które powodują, że w ogólności nie jest wygodna w użyciu. Nie jest ona zadana na całej przestrzeni fazowej a jedynie na ściśle wyznaczonym płacie energetycznym, nie posiada gładkiej gęstości itp. Okazuje się, że można zdefiniować inną miarę, także inwariatną na dynamikę, jednak znacznie wygodniejszą w obliczeniach bo zadaną na całej przestrzeni fazowej i posiadającą gładką gęstość, tak zwaną miarę kanoniczną Gibbsa. Aby jednak sensownie umotywować jej istnienie i zrozumieć jej istotę wygodnie jest posłużyć się pewnym eksperymentem myślowym:

Eksperyment myślowy 4.2.0.1 *Wyobraźmy sobie przestrzeń fazową pewnego układu z której losujemy stany z rozkładu jednostajnego (na całej przestrzeni). Wylosowane stany grupujemy w porcje o jednakowych rozmiarach.Następnie wyliczamy histogram częstości występowania każdego ze stanów w rozważanych grupkach. Histogram ten zadaje pewien empiryczny rozkład prawdopodobieństwa, który jak zaraz pokażemy zbiega do pewnej miary na przestrzeni fazowej.*

Fakt 4.2.0.1 *(Nieformalne sformułowanie twierdzenia Sanowa, patrz np. [11]) Niech $P_e^{(N)}$ oznacza miarę empiryczną rozważaną w 4.2.0.1. Wtedy zachodzi:*

$$\mathbb{P}(P_e^{(N)} \sim P) = e^{-N\mathbb{H}(P)}$$

Prawdopodobieństwo znalezienia się miary empirycznej w pobliżu pewnego rozkładu probabilistycznego P zachowuje się jak $e^{-N\mathbb{H}(P)}$, gdzie $\mathbb{H}(P) = \mathbb{H}(P|\mu)$ jest entropią względną rozkładu P względem miary jednostajnej μ.

Eksperyment myślowy 4.2.0.2 *Nawiązując do eksperymentu myślowego (4.2.0.1), rozważmy teraz sytuację w której wybieramy tylko takie grupki stanów, które mają średnią energię $\overline{E}$, o reszcie zapominamy. Możemy założyć, że wylosowaliśmy tych grupek odpowiednio dużo. Wtedy takie grupki również zadają pewien rozkład empiryczny, który podlega prawu opisanemu w fakcie (4.2.0.1).*

Wniosek 4.2.0.1 *Układ o średniej energii $\overline{E}$ znajdzie się najprawdopodobniej w stanie statystycznym P minimalizującym entropię względną względem[7] rozkładu jednostajnego na przestrzeni fazowej, z zastrzeżeniem że $\int_\Omega Edp = \langle E \rangle_P = \overline{E}$.*

[6] Aby normalizacja była możliwa, miara na całej hiperpowierzchni energetycznej musi być skończona. Dla naszych potrzeb możemy założyć, że prędkości w układzie są ograniczone, wtedy także odpowiednia hiperpowierzchnia jest ograniczona. W ogólności można z takich problemów wybrnąć także w inny sposób, zakładając że prędkości mają rozkład Maxwella.

[7] "entropia względna względem" nie brzmi być może najlepiej stylistycznie, warto jednak podkreślić, że chodzi o entropię względną która różni się od zwykle rozważanej entropii.

Po tych wstępnych faktach jesteśmy gotowi sformułować najważniejsze twierdzenie:

Twierdzenie 4.2.0.2 *(Zasada wariacyjna)*

$$\forall_{\beta>0} - \ln\left(\left\langle e^{-\beta E}\right\rangle_{\mu}\right) = \inf_{P}\left(\mathbb{H}(P) + \beta\langle E\rangle_P\right)$$

i minimum prawej strony jest osiągane przez miarę Gibbsa

$$P_{\beta E}(\sigma) = \frac{e^{-\beta E(\sigma)}}{\langle e^{-\beta E}\rangle_{\mu}}\mu(\sigma)$$

Dowód 4.2.0.3 *Dla dowolnego rozkładu P mamy*

$$\begin{aligned}
\mathbb{H}(P) + \beta\langle E\rangle_P =& \mathbb{H}(P|\mu) + \langle \beta E\rangle_P = \\
=& \left\langle \ln\left(\frac{dP}{d\mu}\right)\right\rangle_P + \langle \beta E\rangle_P = \\
=& \left\langle \ln\left(\frac{dP}{dP_{\beta E}}\right)\right\rangle_P + \left\langle \ln\left(\frac{dP_{\beta E}}{d\mu}\right)\right\rangle_P + \langle \beta E\rangle_P = \\
=& \mathbb{H}(P|P_{\beta E}) + \left\langle \ln\left(\frac{dP_{\beta E}}{d\mu}\right) + \beta E\right\rangle_P = \\
=& \mathbb{H}(P|P_{\beta E}) + \left\langle \ln\left(\frac{\frac{e^{-\beta E}}{\langle e^{-\beta E}\rangle_{\mu}}d\mu}{d\mu}\right) + \beta E\right\rangle_P
\end{aligned}$$

Dalej mamy

$$\begin{aligned}
\mathbb{H}(P|P_{\beta E}) + \left\langle -\beta E + \ln\left(\frac{1}{\langle e^{-\beta E}\rangle_{\mu}}\right) + \beta E\right\rangle_P =& \mathbb{H}(P|P_{\beta E}) + \left\langle -\ln\left(\langle e^{-\beta E}\rangle_{\mu}\right)\right\rangle_P = \\
=& \mathbb{H}(P|P_{\beta E}) - \ln\left(\langle e^{-\beta E}\rangle_{\mu}\right)
\end{aligned}$$

Wystarczy teraz pokazać, że $\mathbb{H}(P|P_{\beta E}) \geq 0$. Korzystając z nierówności $x\ln(x) \geq x - 1$ mamy

$$\begin{aligned}
\mathbb{H}(P|P_{\beta E}) =& \left\langle \frac{P}{P_{\beta E}}\ln\left(\frac{P}{P_{\beta E}}\right)\right\rangle_{P_{\beta E}} \geq \\
\geq& \left\langle \frac{P}{P_{\beta E}} - 1\right\rangle_{P_{\beta E}} = 0
\end{aligned}$$

Ostatecznie wiadomo, że $\mathbb{H}(P|P_{\beta E}) = 0$ wtedy i tylko wtedy, gdy $P = P_{\beta E}$.

□

Twierdzenie 4.2.0.4 *Dla każdej wartości energii $E_{min} < \overline{E} < E_{max}$ istnieje dokładnie jedno takie $\beta = \beta(\overline{E})$, że $\langle E\rangle_{P_{\beta E}} = \overline{E}$.*

Dowód 4.2.0.5 *Twierdzenie 2.4.0.1 z rozdziału 2.4 implikuje, że przy $\beta \to \infty$ miara Gibbsa skupia swoją masę na minimalizatorach energii. Analogiczne rozumowanie pokazuje, że przy $\beta \to 0$ miara Gibbsa skupia masę na maksymalizatorach energii. Ponieważ wartość oczekiwana energii względem miary Gibbsa w sposób ciągły zależy od parametru β, zatem z własności*

Tylko do celów edukacyjnych

Darboux dla funkcji ciągłej wynika, że istnieje takie $\beta = \beta(\overline{E})$, że $\langle E\rangle_{P_{\beta E}} = \overline{E}$. Jedyność β wymaga zauważenia następującego faktu:

$$\frac{\partial}{\partial\beta}\langle E\rangle_{P_{\beta E}} = -Var_{P_{\beta E}}(E) \tag{4.3}$$

Przy założeniu, że energia nie jest stała na całej przestrzeni fazowej (w takim wypadku twierdzenie jest trywialne), wariancja jest ściśle większa od zera, zatem lewa strona (4.3) jest ujemna, zatem średnia energia jest malejącą funkcją β.

□

Twierdzenie 4.2.0.6 *Przy zadanej energii średniej miara Gibbsa minimalizuje entropię względem miary jednostajnej.*

Dowód 4.2.0.7 *Wiemy już z dowodu zasady wariacyjnej, że miara Gibbsa minimalizuje $\mathbb{H}(P) + \beta\langle E\rangle_P$. Weźmy zatem takie $\beta = \beta(\overline{E})$, że $\langle E\rangle_{P_{\beta E}} = \overline{E}$. Mamy wtedy:*

$$\begin{aligned}
\mathbb{H}(P_{\beta E}) + \beta\langle E\rangle_{P_{\beta E}} &= \inf_{P,\langle E\rangle_P=\overline{E}} \mathbb{H}(P) + \beta\langle E\rangle_P \\
&\Updownarrow \\
\mathbb{H}(P_{\beta E}) + \beta\overline{E} &= \inf_{P,\langle E\rangle_P=\overline{E}} \mathbb{H}(P) + \beta\overline{E} \\
&\Updownarrow \\
\mathbb{H}(P_{\beta E}) &= \inf_{P,\langle E\rangle_P=\overline{E}} \mathbb{H}(P)
\end{aligned}$$

□

Miary kanoniczne Gibbsa są z wielu przyczyn bardziej przydatne niż miary mikrokanoniczne, posiadają dobrze zdefiniowaną gęstość i nie ograniczają się jedynie do płatów energetycznych. Warto tu też dodać, że odpowiednia dynamika Glaubera zachowuje miarę Gibbsa w rozważanych przez nas układach spinowych (twierdzenie 3.2.0.11 tego dowodzi) [8].

4.3. Miary graniczne Gibbsa w sensie klasycznym i DLR

Jak wspomniano na początku tego rozdziału, istotą fizyki statystycznej jest badanie własności układów, gdy ich rozmiar zbiega do nieskończoności (granicy termodynamicznej). Pojawiają się tu jednak pewne problemy: dla układu nieskończonego wyrażenia na Hamiltoniany (energie) nie są zbieżne! Nie możemy zatem definiować miar probabilistycznych na przestrzeni fazowej nieskończonego układu w sposób opisywany do tej pory. Okazuje się jednak, że można sobie poradzić w nieco inny sposób. Klasyczna konstrukcja polega na wzięciu dowolnej zmiennej lokalnej F_B na przestrzeni fazowej. Lokalność oznacza, że funkcja ta zależy jedynie od pewnego podzbioru B przestrzeni układu[9], czyli $F_B(\sigma) = F_B(\sigma \cap B), \quad \sigma \in \Omega_\Lambda$. W przypadku modelu gazu klasycznego przejście do granicy termodynamicznej składa się z następujących etapów:

[8] Tak naprawdę dynamika gazu, jak i wiele innych przyzwoitych dynamik tę miarę także zachowuje, nie będziemy tego jednak tutaj dowodzić.

[9] Przestrzeni w której układ jest określony, nie przestrzeni fazowej!

Tylko do celów edukacyjnych

1. Rozważmy wstępujący ciąg podzbiorów $\Lambda_n \nearrow \mathbb{R}^3$

$$\Lambda_1 \subset \Lambda_2 \subset \ldots \Lambda_n \subset \ldots$$

taki, że

$$\bigcup \Lambda_n = \mathbb{R}^3$$

2. Określamy sekwencję ilości cząstek w układzie N_i w ten sposób, że

$$\frac{N_i}{|\Lambda_i|} \to \rho$$

Ustalamy zatem graniczną gęstość cząsteczek na ρ.

3. Jeśli chcemy liczyć średnie względem rozkładów mikrokanonicznych ustalamy sekwencję wartości energii E_i

$$\frac{E_i}{|\Lambda_i|} \to e$$

Wtedy dla każdej zmiennej lokalnej istnieje granica

$$\lim_{\Lambda \nearrow \mathbb{R}^3} \langle F_B \rangle^{micro}_{\Lambda,N,E} = \langle F_B \rangle^{micro}_{\infty,\rho,e}$$

(średnia względem granicznego rozkładu mikrokanonicznego) oraz

$$\lim_{\Lambda \nearrow \mathbb{R}^3} \langle F_B \rangle^{canon}_{\Lambda,N,E} = \langle F_B \rangle^{canon}_{\infty,\rho,e}$$

(średnia względem granicznego rozkładu kanonicznego). Pojawia się pytanie czy powyższe średnie mogą być rzeczywiście interpretowane jako średnie względem pewnych granicznych rozkładów prawdopodobieństwa. Okazuje się, że dla pewnych parametrów tak jest i rzeczywiście można pokazać istnienie granicznego rozkładu. Klasyczna konstrukcja (opisana w czwartym rozdziale [24]) ma jednak pewne wady. Psuje się w pewnych wypadkach w których miara graniczna nie jest wyznaczona jednoznacznie (sytuację w której współistnieje wiele miar granicznych nazywamy przejściem fazowym). Aby dobrze zbadać to zjawisko (będące centralnym zagadnieniem tej pracy) będziemy potrzebować innego narzędzia konstrukcji miar granicznych, mianowicie zaproponowanych przez Dobrushina, Lanforda i Ruelle'a miar DLR. Cała idea polega na nieco innej konstrukcji funkcji energetycznej (przez wzięcie pod wzgląd "zamrożonego" brzegu wokół domeny) i utworzenia miary warunkowej względem takiej funkcji. Następnie przechodzi się z rozmiarem domeny do nieskończoności konstruując miarę na sigma-algebrze zbiorów cylindrycznych (używając miar zdefiniowanych na skończonych cylindrach z zadanym brzegiem). Można pokazać, że miara graniczna w takim sensie zawsze istnieje, co więcej, może istnieć wiele takich miar. Jak później pokażemy przypadek w którym istnieje wiele takich miar jest dla nas szczególnie interesujący. Jeśli istnieje więcej niż jedna miara graniczna to także każda wypukła kombinacja takich miar jest miarą graniczną Gibbsa w sensie DLR. Dla skupienia uwagi przyjmijmy teraz, że będziemy rozważać układ spinowy. Sformalizujmy teraz wypowiedziane wyżej twierdzenia: Rozważmy pewien skończony podzbiór $\Lambda \subset \mathbb{Z}^d$ oraz następujący Hamiltonian:

$$H_\Lambda(\overline{\sigma}) = \sum_{\sigma_i \in \Lambda \vee \sigma_j \in \Lambda, ||\sigma_i - \sigma_j|| \leq R} U(i,j)\sigma_i\sigma_j - h\sum_{\sigma_i \in \Lambda} \sigma_i \tag{4.4}$$

Tylko do celów edukacyjnych

Jak widać[10] Hamiltonian ten różni się od (4.1) o czynniki powstające ze spinów σ nie należących do Λ, a do jego uzupełnienia Λ' (a właściwie warstwy otaczającej Λ o grubości R, gdzie R jest promieniem oddziaływania). O ile tylko promień oddziaływania jest skończony, taka suma jest zawsze zbieżna. Co więcej, (4.4) ma miłą własność iterowania: jeśli $\Lambda_1 \supset \Lambda$ to

$$(H_{\Lambda_1})_\Lambda(\overline{\sigma}) = H_\Lambda(\overline{\sigma})$$

Następnie konstruujemy miary na skończonych podzbiorach $\Lambda \subset \mathbb{Z}^d$, względem takiego nowego Hamiltonianu. Wtedy dla każdej ustalonej konfiguracji zewnętrznej η zadanej na uzupełnieniu kraty Λ^c, powstaje unikatowa skończona miara Gibbsa:

$$\mu_\Lambda^\eta(\sigma_\Lambda) = \frac{1}{Z_{\beta,\Lambda}^\eta} e^{-\beta H_\Lambda(\sigma_\Lambda|\eta_{\Lambda^c})} \tag{4.5}$$

Cała idea teraz polega na tym, że rodzina miar μ_Λ^η powinna być rodziną miar warunkowych dla pewnej miary μ_β. Jak się okazuje miary μ_Λ^η spełniają wszelkie zależności które powinny spełniać miary warunkowe[11]. Dobrushin wpadł na pomysł aby zdefiniować formalne pojęcie następującej postaci:

Fakt 4.3.0.2 *Miara probabilistyczna μ_β na $(\{-1,1\}^{\mathbb{Z}^d}, \mathcal{F})$ jest miarą Gibbsa dla Hamiltonianu H i temperatury odwrotnej β wtedy i tylko wtedy, gdy jej rozkłady warunkowe (ograniczone do skończonego podzbioru Λ kraty, przy ustalonej konfiguracji zewnętrznej) dane są przez (4.5).*

Oczywiście od razu pojawiają się pytania:

- Czy taka miara zawsze istnieje?
- Czy jest wyznaczona jednoznacznie?

Odpowiedź na pierwsze pytanie jest pozytywna - można wykazać istnienie takiej miary. Rozważania do tego konieczne muszą uwzględniać strukturę topologiczną nieskończonej przestrzeni produktowej (wyprodukowanej z przestrzeni wartości przyjmowanych przez pojedynczy spin). Trzeba skonstruować także sigma-algebrę produktową i pokazać, że dla pewnych wstępujących ciągów absorbujących całą przestrzeń $\mathbb{Z}^d$ istnieje zawsze granica miar. Szczegóły tego rozumowania można zobaczyć w [5, 31, 24]. Odpowiedź na drugie pytanie jest natomiast negatywna, jak zobaczymy w niskich temperaturach może istnieć więcej niż jedna miara DLR.

4.4. Struktura zbioru miar DLR, pojęcie przejścia fazowego

Struktura zbioru miar DLR jest jak się okazuje dość bogata. Sformułujemy (bez dowodu) kilka twierdzeń. Odpowiednie dowody (nierzadko bardzo skomplikowane) można zaleźć w [31]. Oznaczmy przez $\mathcal{G}$ zbiór miar DLR. Prawdziwe są następujące twierdzenia:

Twierdzenie 4.4.0.8 *Zbiór $\mathcal{G}$ jest niepusty.*

Twierdzenie 4.4.0.9 *Zbiór $\mathcal{G}$ jest wypukłym i zwartym zbiorem w słabej topologii na zbiorze miar nad $\{-1,1\}^{\mathbb{Z}^d}$.*

[10] Być może nie na pierwszy rzut oka...

[11] Dość łatwo to sprawdzić. Warto tu jednak zwrócić uwagę, że zawdzięczamy to zmodyfikowanemu pojęciu Hamiltonianu.

Tylko do celów edukacyjnych

Twierdzenie 4.4.0.10 *Ekstremalne miary DLR $\mathcal{G}_e$ (takie które nie są kombinacjami wypukłymi innych) mają wzajemnie rozłączne nośniki.*

Oznacza to, że na każdym zbiorze konfiguracji $\{\overline{\sigma}\}$ niezerowym względem pewnej granicznej miary ekstremalnej DLR, wszystkie inne graniczne miary ekstremalne DLR się zerują.

Twierdzenie 4.4.0.11 *Każdy element $\mu \in \mathcal{G}$ może być w sposób jednoznaczny opisany jako całka nad zbiorem $\mathcal{G}_e$ ekstremalnych miar DLR (lub jako kombinacja wypukła, gdy $\mathcal{G}_e$ jest skończony).*

Elementy $\mathcal{G}_e$ nazywamy *czystymi fazami* (*"pure states"* lub *"pure phases"*), w przeciwieństwie do elementów $\mathcal{G} \setminus \mathcal{G}_e$ zwanych *mieszaninami faz* (*fazami mieszanymi*). Widać zatem, że teoria miar DLR jest bogata, a struktura zbioru $\mathcal{G}$ dość ciekawa. Okazuję się, że zbiór $\mathcal{G}$ ulega degeneracji dla odpowiednio wysokiej temperatury β^{-1}:

Fakt 4.4.0.3 *(kryterium wysokotemperaturowe Dobrushina) Dla odpowiednio wysokiej temperatury β^{-1} istnieje dokładnie jedna miara graniczna DLR.*

Dokładny opis tego bardzo eleganckiego kryterium znajduje się w [5]. Intuicyjnie sytuację taką można sobie wyobrazić, gdy niezależnie od zadanych warunków na brzegu dowolnego obszaru skończonego Λ, wewnątrz układu pojawia się praktycznie ta sama miara. Oczywiście istnieją przykłady potencjałów interakcji i ich parametrów, gdy tak nie jest, można jednak zdefiniować dość ogólne warunki, gdy taka sytuacja występuje. Jeśli $H\beta$ jest odpowiednio małe, to miara Gibbsa jest jedynie słabym zaburzeniem miary jednostajnej na przestrzeni fazowej. Układ jest zatem w stanie mocno chaotycznym (każdy stan jest praktycznie tak samo prawdopodobny) i warunki brzegowe nie są w stanie w istotny sposób wpłynąć na zawartość wnętrza Λ. Ciekawym faktem jest to, że istnieje taka temperatura krytyczna która oddziela obszary gdzie może istnieć tylko jedna miara DLR od tych gdzie może współistnieć kilka.

4.5. Argument Peierlsa - przejście fazowe w modelu Isinga

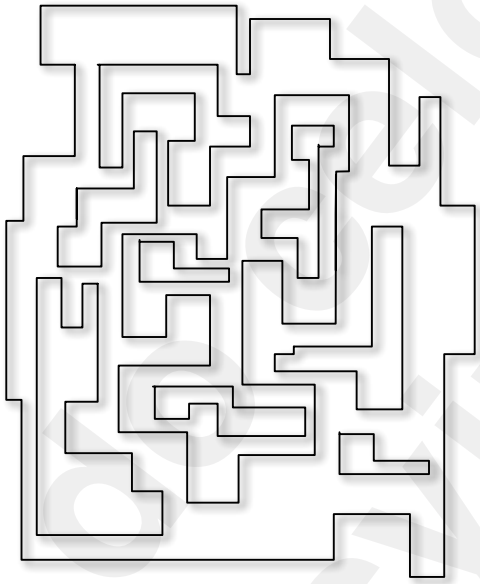
Aby w pełni nacieszyć się bogactwem teorii miar DLR należy wskazać sytuację, gdy miar granicznych jest więcej niż jedna. Wskazanie takiej sytuacji w modelu Isinga oznaczało w praktyce, że model ten dobrze opisuje zjawisko ferromagnetyzmu. Jak już wspominaliśmy Ising w swojej oryginalnej pracy rozważył model jednowymiarowy $d = 1$ w którym przejście fazowe rzeczywiście nie występuje (zawsze jest jedna miara graniczna) i ekstrapolował błędnie swoje rozumowanie na wyższe wymiary. Okazuje się jednak, że już w wymiarze $d = 2$ występują dwie graniczne miary DLR w modelu Isinga, ich konstrukcja wymaga jednak nieco pomysłowości. Pomysł został zaproponowany w [26] w 1936 roku, polega on na obserwacji, że przy dużych β (niskiej temperaturze) miara Gibbsa w modelu Isinga powinna silnie faworyzować stany o niskiej energii. Jeśli pole zewnętrzne $h \neq 0$ to jest tylko jedna taka konfiguracja $\sigma_i = sign(h)$. W sytuacji, gdy $h = 0$ pojawiają się dwa zdegenerowane minima dla $\sigma_i = 1$ i $\sigma_i = -1$. W tym kontekście naturalnym wydaje się scharakteryzowanie stanu opisując jedynie jego rozbieżność od idealnej konfiguracji (zawierającej same 1 lub -1). W ten sposób trafiamy do pojęcia konturu. Rozważmy kratę dualną do $\mathbb{Z}^d$ (kopię kraty $\mathbb{Z}^d$ przesuniętą w każdym wymiarze o $1/2$). Przez $\langle ij \rangle^*$ oznaczamy $d-1$ wymiarową ścianę rozdzielającą punkty kraty σ_i oraz σ_j. Konturem dla konfiguracji $\overline{\sigma}$ nazwiemy:

$$\Gamma(\overline{\sigma}) \equiv \{\langle ij \rangle^* | \sigma_i \sigma_j = -1\}$$

Kontur jest zatem zbiorem krawędzi oddzielających punkty węzłowe kraty $\mathbb{Z}^d$ (spiny) o przeciwnych znakach.

Lemat 4.5.0.12 *Niech Γ będzie konturem zdefiniowanym jak wyżej oraz niech $\partial\Gamma$ będzie jego $d-2$ wymiarowym brzegiem.*

- *$\partial\Gamma(\overline{\sigma}) = \emptyset$ dla każdej konfiguracji $\overline{\sigma}$. Γ może jednak posiadać komponenty nieograniczone.*
- *Niech Γ będzie konturem takim, że $\partial\Gamma(\overline{\sigma}) = \emptyset$. Wtedy istnieją dokładnie dwie takie konfiguracje $\overline{\sigma}$ i $-\overline{\sigma}$ dla których $\Gamma(\overline{\sigma}) = \Gamma(-\overline{\sigma}) = \Gamma$.*



Rysunek 4.1. Przykładowy wygląd fragmentu konfiguracji konturowej.

Źródło: Opracowanie własne

Lemat 4.5.0.13 *Niech Λ będzie skończonym podzbiorem kraty $\mathbb{Z}^d$. Wtedy prawdopodobieństwo wystąpienia w jego wnętrzu konturu Γ wyraża się następująco:*

$$\mathbb{P}(\Gamma) = \frac{e^{-2\beta\left(\sum_{\gamma\in\Gamma}|\gamma|\right)}}{\sum_{\Gamma_\Lambda} e^{-2\beta\left(\sum_{\gamma\in\Gamma_\Lambda}|\gamma|\right)}} = \frac{e^{-2\beta\left(\sum_{\gamma\in\Gamma}|\gamma|\right)}}{Z[\beta]} \tag{4.6}$$

Przez γ oznaczmy komponenty konfiguracji konturowej Γ (pojedyncze spójne składowe).

Dowód 4.5.0.14 *Łatwo sprawdzić, że dla pewnej konfiguracji $\overline{\sigma}$ wewnątrz Λ której odpowiada kontur Γ, zachodzi odpowiedniość:*

$$H_\Lambda(\overline{\sigma}) = 2\sum_{\gamma\in\Gamma}|\gamma| + c(\Lambda) \tag{4.7}$$

Gdzie $c(\Lambda)$ jest zależne od ilości wymiarów (dla ustalonego Λ jest stałe). Teza lematu wynika z definicji miary warunkowej Gibbsa.

□

Tylko do celów edukacyjnych

Lemat 4.5.0.15 *Niech μ_β będzie miarą graniczną Gibbsa dla modelu Isinga przy braku pola zewnętrznego ($h = 0$). Niech γ będzie skończonym elementem konturu. Wtedy:*

$$\mu_\beta(\gamma \in \Gamma(\overline{\sigma})) \leq e^{-2\beta|\gamma|} \tag{4.8}$$

Dowód 4.5.0.16 *Dla każdej konfiguracji konturowej Γ takiej, że $\gamma \in \Gamma$ znajdujemy $\Gamma^* = \Gamma \setminus \{\gamma\}$. Oczywiste jest, że jeśli $\gamma \in \Gamma_1$, $\gamma \in \Gamma_2$ oraz $\Gamma_1^* = \Gamma_2^*$ to $\Gamma_1 = \Gamma_2$. Wtedy z poprzedniego lematu mamy:*

$$\mathbb{P}(\Gamma^*) = e^{2\beta|\gamma|}\mathbb{P}(\Gamma) \tag{4.9}$$

Z czego wynika, że prawdopodobieństwo wystąpienia γ w pewnym Γ szacuje się następująco:

$$\mathbb{P}(\gamma \in \Gamma(\overline{\sigma})) = \sum_{\Gamma \ni \gamma} \mathbb{P}(\Gamma) = e^{-2\beta|\gamma|} \sum_{\Gamma \ni \gamma} \mathbb{P}(\Gamma^*) \leq e^{-2\beta|\gamma|} \tag{4.10}$$

□

Twierdzenie 4.5.0.17 *Niech int_γ oznacza wnętrze elementu konturowego γ. Niech μ_β będzie graniczną miarą Gibbsa dla modelu Isinga z zerowym polem zewnętrznym ($h = 0$). Niech $d \geq 2$. Wtedy istnieje takie $\beta_d < \infty$, że dla $\beta > \beta_d$*

$$\mu_\beta\left(\exists_{\gamma \in \Gamma(\overline{\sigma}) : 0 \in int_\gamma}\right) < \frac{1}{2} \tag{4.11}$$

Notacja $\gamma \in \Gamma(\overline{\sigma}) : 0 \in int_\gamma$ oznacza, że istnieje element konturu, który otacza punkt 0 na kracie $\mathbb{Z}^d$.

Dowód 4.5.0.18 *Mamy, że*

$$\mu_\beta\left(\exists_{\gamma \in \Gamma(\overline{\sigma}) : 0 \in int_\gamma}\right) \leq \sum_{\gamma \in \Gamma(\overline{\sigma}) : 0 \in int_\gamma} \mu_\beta(\gamma \in \Gamma(\overline{\sigma})) \tag{4.12}$$

Zgrubnie szacując ilość konturów γ otaczających 0, takich, że $|\gamma| = k$ otrzymujemy:

$$\sharp\{\gamma \in \Gamma(\overline{\sigma}) : 0 \in int_\gamma, |\gamma| = k\} \leq C_d^k k^{d/(d-1)} \tag{4.13}$$

Gdzie C_d jest stałą zależną od wymiaru. Wtedy otrzymujemy:

$$\mu_\beta\left(\exists_{\gamma \in \Gamma(\overline{\sigma}) : 0 \in int_\gamma}\right) \leq \sum_{k=2d}^{\infty} k^{d/(d-1)e^{-k(2\beta - \ln(C_d))}} \tag{4.14}$$

Biorąc β odpowiednio większe niż $\frac{1}{2}\ln(C_d)$ otrzymujemy żądane oszacowanie.

□

Twierdzenie 4.5.0.19 *W modelu Isinga z zerowym polem zewnętrznym ($h = 0$) dla odpowiednio dużego β istnieją conajmniej dwie graniczne miary Gibbsa w sensie DLR.*

Dowód 4.5.0.20 *Rozważmy wstępujący i absorbujący ciąg zbiorów $\Lambda_n \subset \mathbb{Z}^d$. Rozważmy rodzinę miar skończonych Gibbsa (w sensie DLR) powstałych przy zadaniu na brzegach Λ_n odpowiednio samych 1 lub -1. Nazwijmy te miary odpowiednio $\mu^+_{\beta n}$ i $\mu^-_{\beta n}$. Na podstawie teorii miar DLR wiadomo, że istnieją miary graniczne μ^+_β i μ^-_β. Ale wiemy, że dla odpowiednio dużego β mamy:*

$$\mu^+_\beta(\sigma_0 = -1) \leq \mu^+_\beta\left(\exists_{\gamma \in \Gamma(\overline{\sigma}) : 0 \in int_\gamma}\right) < \frac{1}{2} \tag{4.15}$$

Tylko do celów edukacyjnych

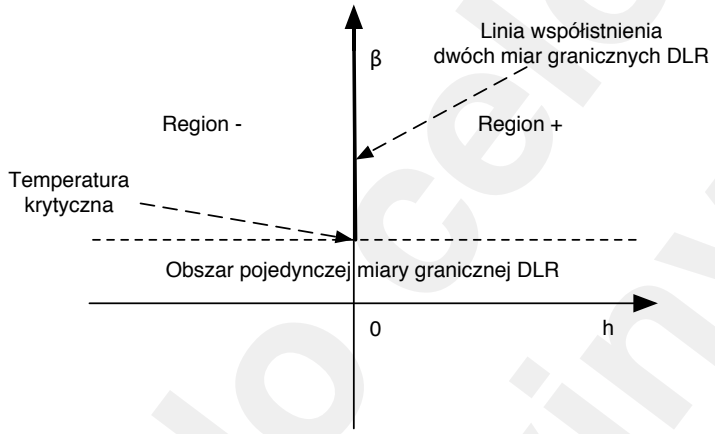
oraz

$$\mu_\beta^-(\sigma_0 = 1) \leq \mu_\beta^-\left(\exists_{\gamma\in\Gamma(\overline{\sigma}):0\in int_\gamma}\right) < \frac{1}{2} \tag{4.16}$$

Zatem

$$\mu_\beta^- \neq \mu_\beta^+ \tag{4.17}$$

□



Rysunek 4.2. Diagram fazowy dla modelu Isinga

Źródło: Opracowanie własne

Twierdzenie (4.5.0.19) jest dość interesującym wynikiem, pokazuje bowiem, że model Isinga rzeczywiście dobrze opisuje ferromagnetyzm. Konstrukcja tego rozumowania jest bardzo elegancka i można ją uogólniać także na inne modele[12].

4.6. Elementy teorii Pirogova-Sinaia

Powróćmy na chwilę do przejścia fazowego w modelu Isinga. Rysunek 4.2 przedstawia diagram fazowy dla tego modelu. Zatrzymajmy się chwilę nad nim i przypatrzmy jakie ma własności. Pozioma oś reprezentuje parametr pola zewnętrznego h, pionowa oś oznacza temperaturę odwrotną β. W tej przestrzeni parametrów istnieje półprosta pokrywająca się z osią β, zaczynająca się w pewnym punkcie krytycznym, na której (dla odpowiednich parametrów β i h) współistnieją dwie miary graniczne DLR. Poniżej punktu krytycznego znajduje się obszar, w którym zawsze istnieje tylko jedna miara graniczna, ze względu wysoką temperaturę. Obszary po bokach linii współistnienia faz także charakteryzują się tym, że istnieje tam tylko jedna miara graniczna DLR, jest to jednak zawsze miara zdominowana przez odpowiedni znak h. Wykreśliliśmy zatem diagram na przestrzeni parametrów Hamiltonianu oraz temperatury odwrotnej. Nasuwa się pytanie, czy w bardziej skomplikowanych przypadkach

[12]W ogólności (gdy układ może przyjmować więcej stanów) stosuje się kontury różnego rodzaju w zależności od tego jakie stany konfiguracji rozdzielają. Rozumowanie podobne do tego w modelu Isinga da się uogólnić na szerszą klasę modeli, co jest jednym z zagadnień teorii Pirogova-Sinaia

Tylko do celów edukacyjnych

można takie diagramy również narysować i jakie będą miały cechy. Zagadnienie to w relatywnie niskich temperaturach rozwiązuje teoria Pirogova-Sinaia, której podstawowe wnioski tu przytoczymy[13].

Na wstępie musimy zauważyć pewne rzeczy: naturalną przestrzenią w której należałoby wyrysować diagram fazowy jest przestrzeń wszystkich Hamiltonianów dobrych dla danego modelu. Oczywiście rysowanie diagramu na przestrzeni funkcyjnej nie jest najbardziej intuicyjne, dlatego rozważa się pewne dobre parametryzacje zbioru dopuszczalnych Hamiltonianów, a następnie rysuje diagram fazowy w odpowiedniej przestrzeni parametrów (tak właśnie postąpiliśmy w przypadku modelu Isinga). W powyższych rozważaniach pojawiają się pewne nieścisłe sformułowania: Co to znaczy, że parametryzacja jest "dobra"? Co to znaczy, że Hamiltonian jest dopuszczalny? Aby rozwiać te wątpliwości potrzebujemy kilku pojęć.

Definicja 4.6.0.1 *Niech $\overline{\sigma}$ będzie konfiguracją na $\mathbb{Z}^d$. $\overline{\sigma}$ nazywamy* stanem bazowym *(*ground state*), jeśli dla dowolnej konfiguracji $\overline{\sigma}'$ różniącej się od $\overline{\sigma}$ na skończonej liczbie pozycji i dla dowolnego ograniczonego $\Lambda \subset \mathbb{Z}^d$ obejmującego wszystkie miejsca gdzie występują różnice zachodzi:*

$$H_\Lambda(\overline{\sigma}) < H_\Lambda(\overline{\sigma}')$$

Stan bazowy jest zatem swego rodzaju minimum energetycznym. Powyższa definicja (podobnie jak niektóre inne rozważania w tym rozdziale) nie jest zbyt ścisła (posiada pewne drobne luki[14]), dokładne rozważania na ten temat można odnaleźć w [35]. Jak pamiętamy, Hamiltonian to funkcja dość specyficznej postaci, w której sumujemy wkłady energetyczne związane z pewnymi potencjałami. W ogólności można rozważać potencjały pomiędzy dowolną ilością cząstek, jednak w naszych rozważaniach przyjęliśmy potencjały zależne jedynie od jednej i dwóch cząstek. Niech teraz H_0 będzie pewnym Hamiltonianem dla układu spinowego, ze zbiorem stanów bazowych G. Możemy teraz rozważyć *perturbację* takiego Hamiltonianu daną następująco:

$$H = H_0 + \sum_{i=1}^{N} \mu_i H_i$$

dla pewnej rodziny parametrów μ_i oraz Hamiltonianów H_i. Wprowadźmy teraz istotną definicję:

Definicja 4.6.0.2 *Niech $(b_1, \ldots, b_r)$ oznacza rodzinę stanów bazowych Hamiltonianu H_0, oraz $t_\mu = (t_\mu(b_1), \ldots, t_\mu(b_r))$, gdzie $t_\mu(b_q) = h_\mu(b_q) - \min_{1\leq m\leq r} h_\mu(b_m)$ (h_μ oznacza specyficzną energię stanu względem Hamiltonianu H_μ). Wtedy powiemy, że perturbacja $H_\mu = H_0 + \sum_{i=1}^{r} \mu_i H_i$ całkowicie rozdziela stany bazowe H_0, jeśli t_μ odwzorowuje przestrzeń parametrów na cały brzeg dodaniego, r-wymiarowego oktantu.*

Parametry μ zwyczajowo nazywa się uogólnionymi polami zewnętrznymi. Załóżmy teraz, że mamy system spinowy z Hamiltonianem H_0 o r stanach bazowych. Weźmy teraz rodzinę sparametryzowanych Hamiltonianów H_μ, gdzie perturbacja H_μ całkowicie rozdziela stany bazowe H_0. Teoria Pirogova-Sinaia implikuje w takim przypadku następujące wnioski (dla pewnego zakresu niskich temperatur):

- W przestrzeni parametrów μ istnieje pojedynczy punkt (*punkt współistnienia faz*), dla którego w wyjściowym układzie współistnieje r miar granicznych DLR.

[13] W całej okazałości teoria ta jest bardzo skomplikowana technicznie i jej przytoczenie daleko wykraczałoby poza zakres tej pracy.

[14] Dokładne definicje potrzebne do rozważania tych pojęć wychodzą poza ramy tej pracy, a w tym kontekście nie są niezbędnie potrzebne.

- Dla każdego zbioru $\pi \subset \{1, \ldots, r\}$ mocy $r-1$ istnieje krzywa gładka na której współistnieje $r-1$ miar granicznych DLR odpowiadających wybranym stanom bazowym. Wszystkie krzywe tego typu spotykają się w punkcie współistnienia wszystkich miar granicznych DLR.

- W ogólności dla każdego zbioru $\pi \subset \{1, \ldots, r\}$ mocy k istnieje $r-k$ wymiarowa hiperpowierzchnia gładka na której współistnieje k miar granicznych DLR odpowiadających wybranym stanom bazowym. Powierzchnie odpowiadające zbiorom π i π' spotykają się na odpowiedniej hiperpowierzchni wyznaczającej współistnienie miar dla zbioru indeksów $\pi \cup \pi'$.

Należy tu mocno podkreślić, że w ogólności wyniki teorii Pirogova-Sinaia prawdziwe są tylko w pewnym otoczeniu zerowej temperatury i jedynie w pewnych modelach (np. w modelu Isinga) wiadomo jak wygląda cały diagram fazowy. Można przypuszczać, że wnioski teorii Pirogova-Sinaia są zachowane do samej temperatury krytycznej (w której wszystkie fazy zlewają się w jedną), nie ma jednak formalnych dowodów potwierdzających tego typu dywagacji, nie ma także w tej chwili istotnych kontrprzykładów.

Tylko do celów edukacyjnych

ROZDZIAŁ 5

Model lokalnie Hopfieldowskiej sieci neuronowej

Centralnym zagadnieniem tej pracy, jest numeryczny eksperyment, który potwierdza, że teoria Pirogova-Sinaia dobrze opisuje lokalnie Hopfieldowską sieć neuronową. Wiadomo, że teoria ta dobrze opisuje podobne układy gdy ich rozmiar zbiega do nieskończoności co jednak wcale nie implikuje, że w układach skończonych będzie można zaobserwować jakąkolwiek zgodność z przewidywaniami teoretycznymi. Okazuje się, że już w całkiem niewielkich układach pojawiają się regularności i efekty doskonale pasujące do teorii Piorogova-Sinaia.

5.1. Pojęcia wstępne

Badania opisane w tej pracy, zostały przeprowadzone na *lokalnie Hopfieldowskiej sieci neuronowej*. Formalnie taką sieć można opisać w następujący sposób:

Definicja 5.1.0.3 *Sieć neuronową nazwiemy lokalnie Hopfieldowską jeśli:*

- *Istnieje na niej lokalna struktura połączeń, to znaczy sieć ma w przybliżeniu strukturę kraty d wymiarowej, każdy neuron kontaktuje się jedynie z pewną ilością najbliższych sąsiadów.*
- *Sieć uczona jest za pomocą reguły Hebba wzorców i tworzy autoasocjator.*

Oczywiście z dowolnej sieci Hopfielda łatwo zrobić sieć lokalnie Hopfieldowską poprzez zlikwidowanie dalekich połączeń.

Definicja 5.1.0.4 *Sieć lokalnie Hopfieldowską nazwiemy jednorodną jeśli struktura połączeń lokalnych w każdym miejscu jest taka sama.*

W toku dalszych rozważań będziemy się zajmować sieciami jednorodnymi w sensie jak powyżej.

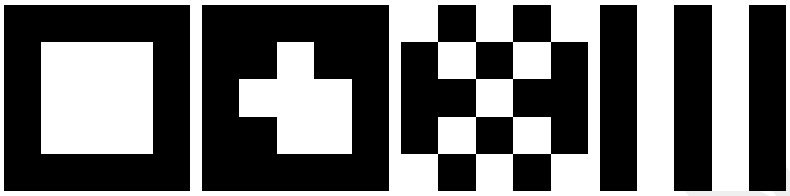
5.2. Opis modelu

W symulacjach użyty został układ składający się z siatki 250×250 (62500) neuronów każdy przyjmujący stany ze zbioru $\{-1, 1\}$. Każdy neuron jest połączony ze swoim kołowym otoczeniem o promieniu 4 neuronów. Taka sieć uczona jest zestawu cyklicznie powtarzanych wzorców 5×5. Każda komórka obejmująca 5×5 neuronów odpowiadająca podstawowemu wzorcowi będzie dalej określana *komórką podstawową*. Diagramy użyte w symulacji są zaprezentowane na obrazku (5.1). Każdy wzorzec wchodzi do wag sieci w parametrem α_i:

$$w_{i,j} = \sum_{k=1}^{N} \alpha_k P_k[i] P_k[j]$$

przy czym

$$\sum_{k=1}^{N} \alpha_k = 1$$



Rysunek 5.1. Przykłady skorelowanych wzorców użytych w symulacji. Najbardziej lewy ($\mathcal{P}_1$) jest wzorcem silnie dominującym (na dalszych diagramach fazowych będzie oznaczany kolorem czerwonym). Drugi w kolejności z lewej ($\mathcal{P}_2$, oznaczany kolorem zielonym na diagramach) jest silnie skorelowany z pierwszym, razem tworzą "koalicję" wspomagając się przeciw trzeciemu wzorcowi ($\mathcal{P}_3$, oznaczany niebieskim kolorem na diagramach), który ma najmniejszą korelację z innymi. Czwarty wzorzec ($\mathcal{P}_4$) ma również niewielką korelację z innymi.

Źródło: Opracowanie własne

Przypomnijmy, że w tradycyjnej w regule Hebba (2.1) wzorce wchodziły z takimi samymi wagami $\frac{1}{N}$ (gdzie N jest ilością wzorców). W tradycyjnym modelu rozważaliśmy jednak wzorce niezależne, nieskorelowane w przeciwieństwie do badań przedstawionych w tej pracy, gdzie wzajemne nietrywialne korelacje wzorców są dla nas szczególnie interesujące[1]. Korelacje te (jak zobaczymy już niedługo) będą miały istotny wpływ na kształt diagramu fazowego (jak się okaże dalej, zestaw parametrów α będzie odpowiedzialny za parametryzację Hamiltonianu, zatem będzie jednoznacznie determinował fazę układu.). Sieć posiada ustalony warunek brzegowy dany przez jeden z rozważanych wzorców. Każda symulacja składa się z następujących etapów:

- Ustalenia zestawu parametrów α_i.
- Ustalenia wag sieci.
- Ustalenia temperatury odwrotnej β.
- Ustalenia numeru wzorca wstawionego jako warunek brzegowy.
- Losowego zainicjowania stanu początkowego sieci.
- Doprowadzenia sieci do stanu przybliżonej równowagi termodynamicznej (rozkładu stacjonarnego).

Po wykonaniu powyższych etapów możemy stwierdzić, czy wzorzec jest stabilny. Przez stabilność rozumiemy sytuację w której wzorzec zadany na brzegu, zreprodukuje się we wnętrzu swobodnie ewoluującej sieci (brzeg specyfikuje w pełni stan końcowy sieci). O ile teraz zgodzimy się[2], że parametry α dobrze parametryzują Hamiltonian, to jak pamiętamy, teoria Pirogova-Sinaia postuluje, że cały N wymiarowy simplex $\alpha_i \in [0,1]$, $\sum_i \alpha_i = 1$ można podzielić na następujące fragmenty:

- $N-1$ wymiarowe obszary w których dokładnie jeden wzorzec jest stabilny.
- $N-2$ wymiarowe hiperpowierzchnie na których dokładnie dwa wzorce są stabilne.

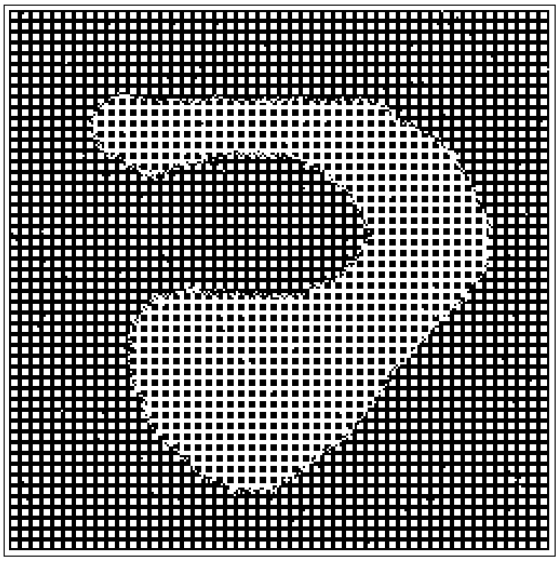
[1] Trzeba to mocno podkreślić - korelacje mają tu kluczowe znaczenie.
[2] Odpowiednia motywacja znajduje się w szczegółach modelu.

Tylko do celów edukacyjnych

- $N - R$ wymiarowe hiperopwierzchnie na których dokładnie R wzorców jest stabilnych.
- Pojedynczy punkt w którym wszystkie wzorce są stabilne (*punkt współistnienia faz*).

Szczegóły modelu wypisano poniżej:

1. Dynamika: We wszystkich symulacjach używana była asynchroniczna dynamika Glaubera.
2. Wyznaczanie stabilnych wzorców: W chwili uzyskania przybliżonej równowagi termodynamicznej trzeba wyznaczyć, który z użytych wzorców najbardziej przypomina stan sieci. W tym celu użyta została procedura wyznaczająca stopień podobieństwa stanu sieci do kolejnych wzorców. Należy tu wspomnieć, że w chwili gdy wzorzec P_i jest stabilny, także jego negatyw wykazuje dużą stabilność (wynika to z degeneracji negatywowej sieci Hopfielda). Metoda wyznaczania współczynnika podobieństwa musi to brać pod uwagę. Uznajemy bowiem, że negatyw wzorca jest równie dobry jak sam wzorzec - wynika to z faktu, że często przez długi czas utrzymywał się następujący stan sieci: pole jednego z wzorców oraz plama jego negatywu gdzieś we wnętrzu. Plamy takie kurczą się i zawsze znikają, jednak niezwykle powoli. Aby oszczędzić czas obliczeń przyjmiemy, że stan końcowy takiej sieci jest już całkowicie zdeterminowany - wiadomo że plama negatywu zniknie i pozostanie sam wzorzec. Dlatego procedura ustalania stabilnego wzorca jest nieczuła na negatywy, w szczegółach wygląda ona następująco:



Rysunek 5.2. Typowa sytuacja występująca w czasie symulacji. Jeden silny wzorzec ($\mathcal{P}_1$) oraz plama jego negatywu w środku. Warto zauważyć, że algorytm zliczający pozycje niezgodności w komórkach podstawowych, zliczy jedynie pozycje w komórkach zawierających zarówno wzorzec jak i negatyw (których jest relatywnie bardzo mało), co skutkuje tym, że parametr D_{glob} będzie bardzo niewielki.

Źródło: Opracowanie własne

- Poniższe czynności wykonujemy po kolei dla każdego wzorca:

- Dla każdej komórki podstawowej liczona jest ilość pozycji na których nie zgadza się ona z aktualnie rozważanym wzorcem.
- Jeśli ilość ta przekracza połowę wszystkich neuronów komórki, oznacza to, że jej stan bardziej przypomina negatyw wzorca i odpowiednio liczymy pozycje niezgodności z negatywem.
- Wyliczamy współczynnik niezgodności komórki według wzorcu

$$D_{kom} = \frac{\#\text{ ilość pozycji niezgodności}}{\#\text{ ilość neuronów w komórce podstawowej}}$$

- Uśredniając D_{kom} po całej dziedzinie obliczeniowej uzyskujemy globalny parametr podobieństwa D_{glob}.
- Jako najbardziej odpowiadający stanowi sieci uznajemy wzorzec którego parametr D_{glob} jest najmniejszy.

Trzeba tutaj zaznaczy, że powyższy sposób działa jedynie w niskich temperaturach gdy odtwarzane wzorce nie są zbyt zaszumione. Ponieważ jednak celem tych badań jest potwierdzenie zasadności opisu tych układów za pomocą teorii Pirogova-Sinaia, więc rozważane temperatury także będą relatywnie niskie.

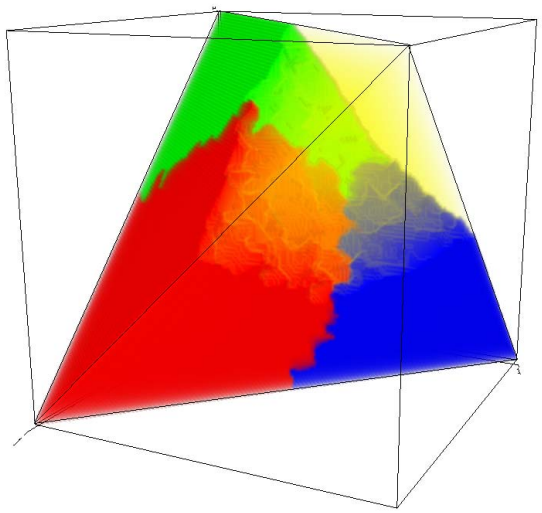
3. Uzyskanie rozkładu stacjonarnego: System, który zawiera 62500 jednostek może być raczej uznany za duży (chyba że posiada się bardzo dużą moc obliczeniową), w związku z tym uzyskanie rozkładu stacjonarnego nie jest zagadnieniem trywialnym. Użycie algorytmu doskonałego Proppa-Wilsona niestety można wykluczyć - przy obecnej prędkości komputerów prawdopodobnie jeden przebieg tego algorytmu trwałby wiele dni (może nawet tygodni). Sytuację ratuje jednak pewna okoliczność: sieć dość szybko osiąga stan taki jak na rysunku (5.2), a w tym przypadku dalsza ewolucja układu jest już praktycznie zdeterminowana. Wiadomo, że wszelkie plamy negatywu wzorca prędzej lub później znikną a dalsza ewolucja nie przyniesie już nic nowego. Warto tutaj zauważyć, że taka sytuacja taka pojawia się zarówno w przypadkach gdy stan sieci zgadza się z warunkiem brzegowym jak i w sytuacji gdy jest inaczej (w takim przypadku pojawia się dodatkowo cienka obwódka na styku różniących się wzorców). Wszystkie te zjawiska powodują, że jak tylko sieć wytworzy pole jednego wzorca (plus plamy jego negatywu)[3], już możemy z bardzo dużą pewnością przewidzieć jak będzie wyglądać stan stacjonarny i zatrzymać sieć.

4. Diagram fazowy: Diagram fazowy został wyrysowany w przestrzeni parametrów $\alpha_1 \dots \alpha_n$. Aby rozważania zawarte w tej pracy miały sens, powyższa parametryzacja Hamiltonianu musi spełniać warunki stawiane przez teorię Pirogova-Sinaia. W praktyce oznacza to, że odpowiednie mapowanie *energii specyficznych* dla stanów bazowych (w tym przypadku stanów wypełnionych wszędzie tym samym wzorcem bazowym) musi wypełniać odpowiedni brzeg n wymiarowego oktantu. W naszym przypadku oznacza to, że wektory złożone z energii specyficznych[4]:

$$\begin{pmatrix} E(P_1|1,0,...,0) \\ E(P_1|0,1,...,0) \\ \vdots \\ E(P_1|0,0,...,1) \end{pmatrix}, \begin{pmatrix} E(P_2|1,0,...,0) \\ E(P_2|0,1,...,0) \\ \vdots \\ E(P_2|0,0,...,1) \end{pmatrix}, \dots, \begin{pmatrix} E(P_n|1,0,...,0) \\ E(P_n|0,1,...,0) \\ \vdots \\ E(P_n|0,0,...,1) \end{pmatrix}$$

[3]Co w toku symulacji następowało zazwyczaj po mniej niż 3 mln kroków.
[4]Zapis postaci $E(P_1|1,0,...,0)$ jest wyjaśniony poniżej.

muszą być liniowo niezależne. Warunek ten zachodzi gdy wzorce $P_1, \ldots, P_n$ są liniowo niezależne, co oczywiście ma miejsce w tym przypadku. Aby zebrać odpowiednią ilość danych do wyrysowania diagramu fazowego wykorzystana została metoda Monte-Carlo. Punkty z simpleksu parametrów α były losowane z rozkładu jednostajnego, następnie dla każdego wylosowanego zestawu parametrów wykonywana była symulacja i określony został stabilny wzorzec. Ten proces był powtarzany około 4000 razy dla każdego diagramu w celu uzyskania zadowalającej rozdzielczości.



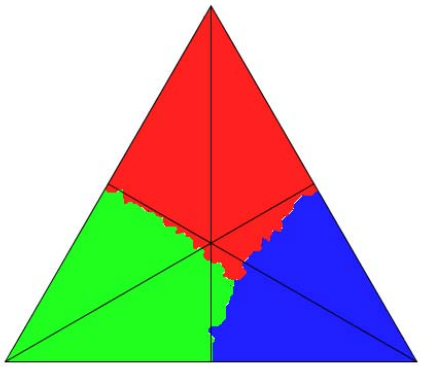
Rysunek 5.3. Przykład diagramu fazowego dla czterech wzorców ($\mathcal{P}_1$,$\mathcal{P}_2$,$\mathcal{P}_3$,$\mathcal{P}_4$,$\beta = 0.5$). Przedni fragment wykresu jest przezroczysty dla lepszego wyeksponowania punktu współistnienia faz. Obszary stabilności pojedynczych wzorców są dość regularne, a ich granice dość ostre.

Źródło: Opracowanie własne

5. Wizualizacja: Wizualizacja objęła przypadki 3 i 4 wzorców, zatem obrazowanymi simpleksami są trójkąt i czworościan. W przypadkach większej ilości wymiarów (wzorców) pojawiają się pewne problemy. Przez pewien czas rozważana była wersja rzutowania barycentrycznego (opisanego w [29]). Niestety tego typu rzutowania szczególnie mocno deformują okolice środka ciężkości rzutowanej figury (do tego stopnia, że rzut przestaje być jednoznaczny), a z punktu widzenia tej pracy właśnie te okolice są najbardziej ciekawe. Ponieważ przestrzeń parametrów była próbkowana losowo metodą Monte-Carlo, użyty został schemat uzupełniający luki w diagramie. Schemat ten sprowadza się do wielokrotnej procedury określającej dla danego punktu wykresu najczęściej występujący kolor jego sąsiadów. Po kilku iteracjach takiej procedury wykres jest całkowicie uzupełniony.

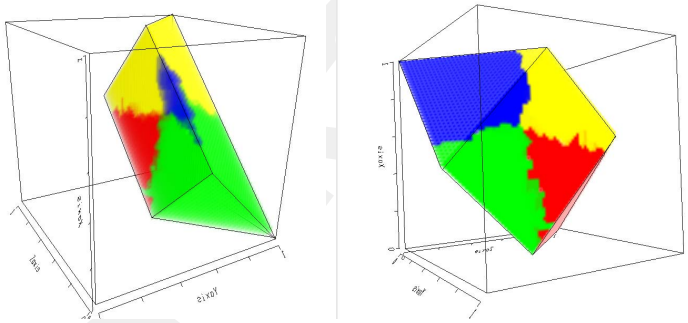
5.3. Wyniki obliczeń i ich związek z przewidywaniami teoretycznymi

W czasie symulacji można było zaobserwować szereg ciekawych zjawisk. To co przykuwa od razu uwagę to regularny kształt diagramu fazowego dobrze pasujący do przewidywań teore-



Rysunek 5.4. Przykład diagramu fazowego dla trzech wzorców ($\mathcal{P}_1$,$\mathcal{P}_2$,$\mathcal{P}_3$). Każde pole koloru oznacza obszar gdzie dokładnie jeden wzorzec jest stabilny. Miejsce w którym wszystkie pola się stykają jest punktem współistnienia faz.

Źródło: Opracowanie własne



Rysunek 5.5. Przykładowe cięcia przez diagram fazowy (cztery wzorce $\mathcal{P}_1$,$\mathcal{P}_2$,$\mathcal{P}_3$,$\mathcal{P}_4$ z rysunku 5.1, $\beta = 0.5$). Widać na nich wyraźnie położenie punktu współistnienia faz, geometrię płaszczyzn współistnienia dwóch faz i krzywych współistnienia trzech faz. Nieregularności granic międzyfazowych to pozostałość po próbkowaniu Monte-Carlo i niedoskonałości tej metody. Generalnie im więcej próbek zebrano tym gładsze były granice.

Źródło: Opracowanie własne

tycznych postulowanych przez teorię Pirogova-Sinaia. Oczywiście eksperyment numeryczny nie może potwierdzić gładkości powierzchni międzyfazowych, można jedynie powiedzieć, że w obrębie rozdzielczości dostępnych dla metody Monte-Carlo nie da się stwierdzić żadnych poważnych nieregularności. Jednym z celów symulacji było zbadanie czy rzeczywiste położenie punktu współistnienia faz nawiązuje do przewidywań zerowo-temperaturowych. Generalnie wyznaczenie położenia punktu wielofazowego jest bardzo trudne[5]. W szczególnym przypadku zerowej temperatury ($\beta^{-1} = 0$) można jednak dość łatwo przyrównać energie warunkowe dla poszczególnych wzorców i uzyskać teoretyczne położenie punktu równowagi. Pewne intuicje każą myśleć, że sprawa się znacznie komplikuje w niezerowych temperaturach, w których trzeba by było porównywać energie swobodne. Intuicje te potwierdza przedstawiony w tej pracy eksperyment obliczeniowy. Wróćmy jednak do przypadku zerowej temperatury. Mamy w takim przypadku do rozwiązania następujący układ równań:

$$\begin{cases} E(P_1|\alpha_1, \alpha_2, ..., \alpha_n) & = E(P_2|\alpha_1, \alpha_2, ..., \alpha_n) \\ E(P_2|\alpha_1, \alpha_2, ..., \alpha_n) & = E(P_3|\alpha_1, \alpha_2, ..., \alpha_n) \\ & \vdots \\ E(P_{n-1}|\alpha_1, \alpha_2, ..., \alpha_n) & = E(P_n|\alpha_1, \alpha_2, ..., \alpha_n) \\ \alpha_1 + \alpha_2 + ... + \alpha_n & = 1, \end{cases} \tag{5.1}$$

gdzie $E(P_x|\alpha_1, ..., \alpha_n)$ to energia wzorca P_x dla układu zainicjowanego parametrami $\alpha_1, ..., \alpha_n$ (energia specyficzna wzorca względem odpowiednio sparametryzowanego Hamiltonianu). Jak pamiętamy energia zdefiniowana jest następująco[6]:

$$\begin{aligned} E(\bar{\sigma}) &= -\frac{1}{2}\sum_{i,j}\sigma_i\sigma_j w_{i,j} = \\ &= -\frac{1}{2}\sum_{i,j}\sigma_i\sigma_j \sum_{k=1}^{n}\alpha_k P_k[i]P_k[j], \end{aligned}$$

Zatem wyrażenia w równaniu (5.1) mogą być rozpisane następująco:

$$\begin{aligned} &E(P_x|\alpha_1, ..., \alpha_n) = \\ &= \alpha_1 E(P_x|1, 0, ..., 0) + ... + \alpha_n E(P_x|0, ..., 1) \end{aligned}$$

W związku z tym 5.1 jest układem równań liniowych dla którego można łatwo znaleźć rozwiązanie. Eksperyment obliczeniowy potwierdził, że dla niskich temperatur (tak niskich, że dobrze przybliżających model zerowo-temperaturowy) punkt współistnienia faz rzeczywiście znajduje się w okolicach rozwiązania (5.1). Wynik ten dla wzorców $\mathcal{P}_1$,$\mathcal{P}_2$,$\mathcal{P}_3$,$\mathcal{P}_4$ prezentuje tabela 5.1. Ponieważ diagram fazowy mamy jedynie w wersji próbkowanej Monte-Carlo, do wyznaczenia położenia punktu współistnienia faz użyto następującego algorytmu aproksymacyjnego:

1. Zainicjuj $x = 1$
2. $x = x \cdot 1/2$

[5] Jest to jeden z otwartych problemów tego zagadnienia, omówionych szerzej w następnym rozdziale.

[6] Dla precyzji trzeba by powiedzieć więcej na temat tego gdzie przebiegają indeksy i, j, jak pamiętamy z poprzednich rozdziałów kształt Hamiltonianu musi być odpowiedni aby rozważać miary DLR. Tutaj jednak nie ma to znaczenia, chodzi o sam kształt funkcji energetycznej.

Tylko do celów edukacyjnych

Tabela 5.1. Porównanie danych teoretycznych i obliczeniowych

	α_1	α_2	α_3	α_4
Wynik teoretyczny (rozwiązanie (5.1))	0.1063	0.2538	0.3268	0.3129
Wynik obliczeniowy	0.1092	0.2588	0.3212	0.3106
Różnica	-0.0029	-0.0050	0.0055	0.0023

3. Przypasuj (losowo lub gradientowo) do diagramu fazowego kostkę o boku x, tak aby było w niej mniej więcej tyle samo próbek należących do każdego wzorca.

4. Jeśli x jest zadowalająco małe (ale na tyle duże, że w kostce o promieniu x mieści się przynajmniej po jednym przedstawicielu każdego wzorca), to zakończ algorytm. W przeciwnym przypadku przejdź do punktu 2, ograniczając obszar poszukiwań do objętości obecnie rozważanej kostki.

Po zrealizowaniu kilkunastu przebiegów powyższego algorytmu dla pojedynczego diagramu i uśrednienieniu wyników, uzyskamy oszacowanie położenia punktu wielofazowego, tak dokładne na ile to pozwala rozdzielczość próbkowania przestrzeni parametrów (diagramu). Jak widać w tabeli 5.1 w niskich temperaturach oszacowanie teoretyczne wynikające z przyrównania energii specyficznych poszczególnych wzorców działa doskonale. Sytuacja zmienia się w wyższych temperaturach, w których istotną rolę zaczyna odgrywać entropia. W pewnym sensie można się było spodziewać, że wraz ze wzrostem temperatury, pojawią się pewne zaburzenia, badania teoretyczne tych zjawisk są narazie trudne (prace w tej dziedzinie trwają, teoria Pirogova-Sinaia zapewnia opis teoretyczny, jest on jednak bardzo skomplikowany i nieefektywny obliczeniowo). Eksperyment przedstawiony w tej pracy ukazuje pewne ciekawe efekty. Jak zostało to uwidocznione w tabeli 5.2, współrzędne punktu współistnienia faz w dość regularny sposób zależą od temperatury odwrotnej β. Eksperyment numeryczny zda-

Tabela 5.2. Ruch punktu wielofazowego (rezultaty obliczeń)

Temp. odwrotna β	α_1	α_2	α_3	α_4
1.0	0.109	0.258	0.321	0.310
0.5	0.117	0.254	0.318	0.308
0.2	0.115	0.259	0.325	0.299
0.1	0.083	0.270	0.338	0.308
0.05	0.047	0.287	0.346	0.319

je się wskazywać, że wraz ze wzrostem temperatury (spadkiem β) punkt współistnienia faz przemieszcza się w ten sposób, że niewielkie parametry α stają się jeszcze mniejsze, zaś te duże stają się jeszcze większe. Odpowiada to sytuacji w której wzorce z dużym udziałem w energii całkowitej zdobywają jeszcze bardziej znaczący udział, na rzecz tych mniej istotnych.

Tylko do celów edukacyjnych

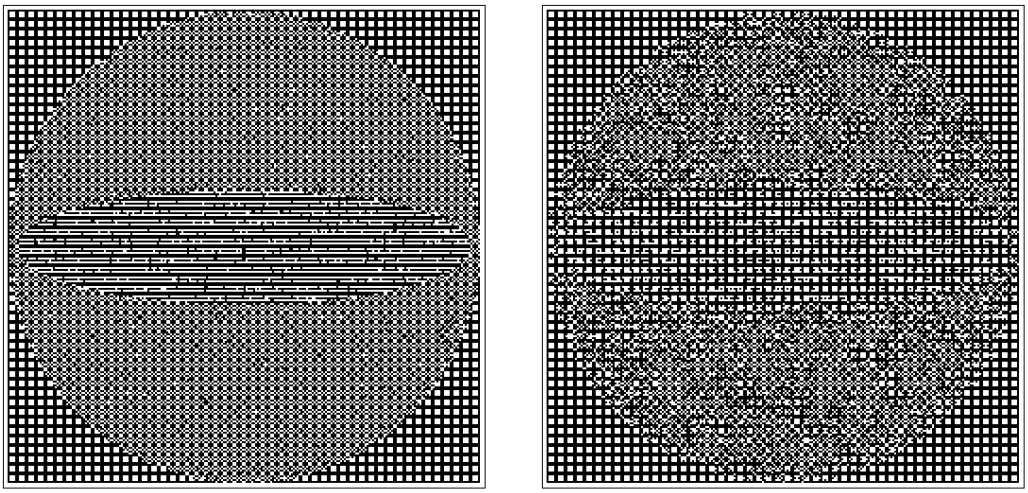
Widać to także w tabeli 5.3, demonstrującej procentowy udział w objętości diagramu czystych faz odpowiadających kolejnym wzorcom. Jak widać procentowy udział czystych faz o największej objętości zwiększa się wraz z temperaturą, tych najmniejszych dalej zmniejsza. Cały proces wydaje się być dość regularny, tak jakby za całym zjawiskiem ukrywała się prosta, monotoniczna zależność. W tej chwili trudno jednak podać konkretny wzór opisujący w sposób analityczny zaobserwowane tutaj zjawisko. Ciekawe zachowania można zaobserwować

Tabela 5.3. Objętości (w %) obszarów czystych faz w diagramach fazowych (rezultat obliczeń).

Temp. odwrotna β	Wzorzec $\mathcal{P}_1$	Wzorzec $\mathcal{P}_2$	Wzorzec $\mathcal{P}_3$	Wzorzec $\mathcal{P}_4$
1.0	33.11	24.84	20.80	21.26
0.5	33.72	24.33	19.91	22.02
0.2	33.41	24.38	20.55	21.63
0.1	34.49	24.43	19.48	21.65
0.05	34.14	26.44	19.34	20.26
0.04	35.47	25.54	19.31	20.59

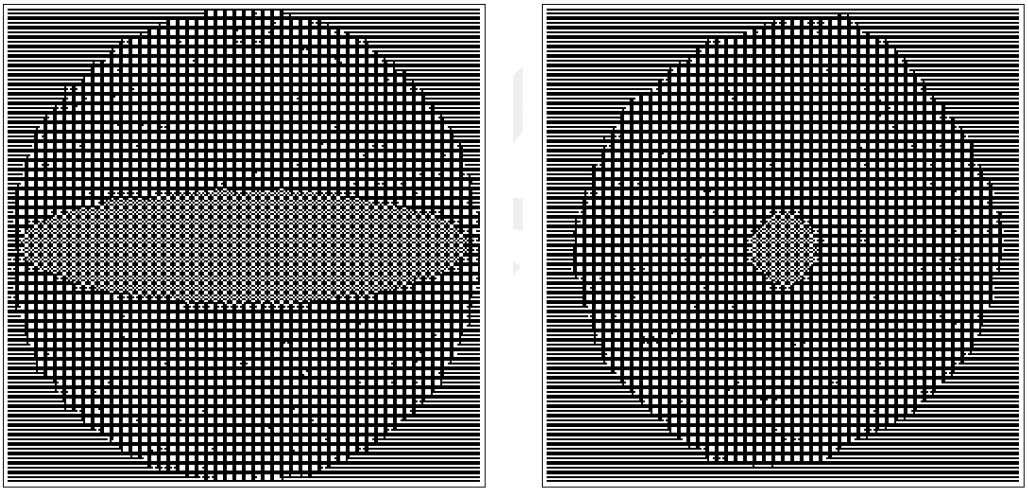
także w samej dynamice symulowanego układu neuronowego (spinowego). Daje się wyróżnić dwie klasy zjawisk - objętościowe i brzegowe. Zjawiska objętościowego pochłaniania jednego wzorca przez drugi występują w przypadku silnej dominacji jednego ze wzorców (parametry α Hamiltonianu z obszaru czystej fazy dla danego wzorca). Są one niezwykle szybkie, duża plama niestabilnego wzorca znika już po kilkudziesięciu tysiącach iteracji (Rysunek 5.6), przy czym proces zanikania przebiega z równą intensywnością wewnątrz jak i na zewnątrz plamy niestabilnego wzorca. Zupełnie inaczej sprawy wyglądają w sytuacji współistnienia wielu faz, kiedy wszystkie wzorce są stabilne. W takim przypadku pojawiają się zjawiska brzegowe związane z napięciami powierzchniowymi (na powierzchniach rozdzielających obszary wypełnione jednym ze wzorców). Jak wiadomo w przestrzeni parametrow istnieje tylko jeden punkt w którym współistnieją wszystkie fazy, a w dodatku w ogólności nie wiadomo jak wyznaczyć jego położenie, w symulacji zatem posługujemy się wynikiem empirycznym dla danej temperatury w nadziei, że przybliża on dobrze położenie rzeczywistego punktu współistnienia faz. Rysunek 5.7 pokazuje ewolucję układu w okolicach punkty współistnienia faz. Aby dobrze zaobserwować co się dzieje w sytuacji stabilnego współistnienia wielu faz, można posłużyć się pewną sztuczką. W naszym układzie nie ma pola zewnętrznego, zatem wykazuje on podobną do modelu Isinga z $h = 0$ własność stabilności fazy dodatniej i ujemnej. Zatem jeśli pewien wzorzec w naszym modelu jest stabilny, to także jego negatyw jest stabilny. W ogólności określenie znaku na brzegu dziedziny obliczeniowej łamię tę symetrię, jednak przez pewien czas można obserwować zjawiska występujące między dwoma idealnie stabilnymi fazami. Rysunek 5.8 pokazuje przykład ewolucji układu w którym jeden wzorzec współistnieje stabilnie ze swoim negatywem. Ruch granic rozdzielających fazy ma charakter ruchu proporcjonalnego do średniej krzywizny[7], co jest charakterystyczne między innymi dla układów z oddziaływaniami zadanymi przez tzw. Potencjały Kaca.

[7]Ponieważ badamy układ dwuwymiarowy to jest tylko jeden parametr krzywizny, zatem sformułowanie "ruch według średniej krzywizny" można tu rozumieć jako "ruch według krzywizny".



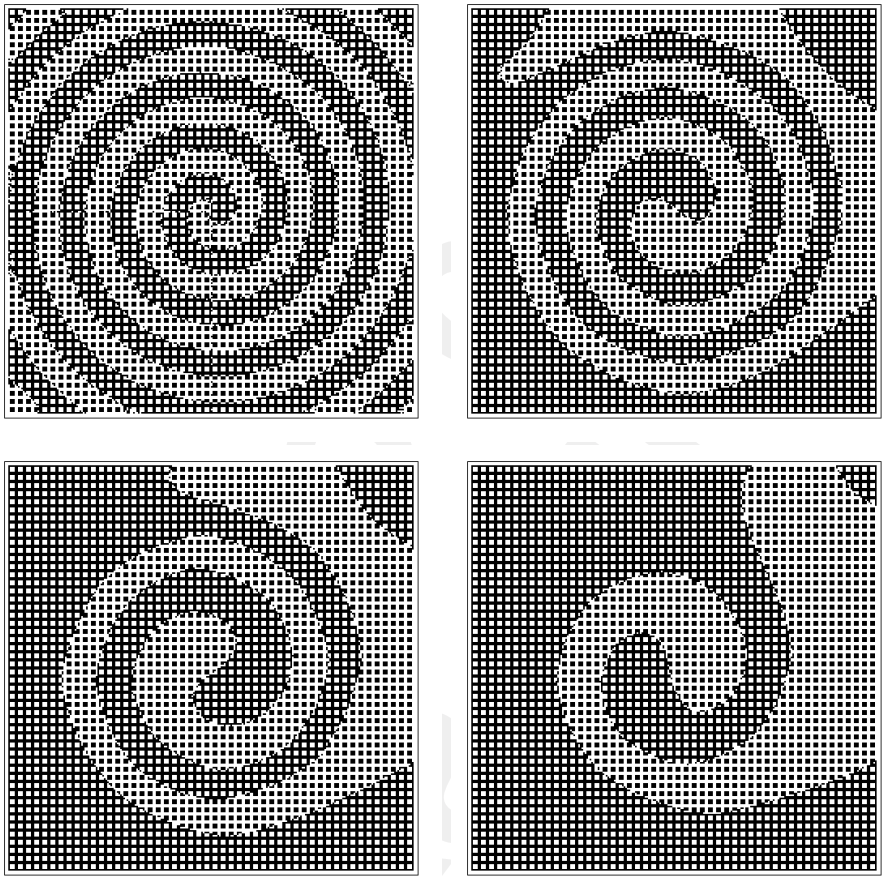
Rysunek 5.6. Ewolucja sieci w warunkach ścisłej dominacji jednego ze wzorców ($\mathcal{P}_1$) w średniej temperaturze ($\beta = 0.2$), w początkowej fazie symulacji (po lewej) oraz po $9 \cdot 10^4$ krokach. Elipsa i kółko utworzone z niestabilnych wzorców ($\mathcal{P}_3$,$\mathcal{P}_4$) są gwałtownie pochłaniane przez stabilny wzorzec. Proces ten jest szybki i gwałtowny, w przeciwieństwie do powolnych procesów brzegowych pokazanych na rysunku 5.7.

Źródło: Opracowanie własne



Rysunek 5.7. Stan sieci w warunkach współistnienia faz w niskiej temperaturze ($\beta = 1$), stan początkowy (na lewo) i po $3.39 \cdot 10^7$ krokach (na prawo). Elipsa i kółko utworzone ze wzorców zwężają się według średniej krzywizny. Proces ten jest powolny i spokojny, gdyż naddatek energii konfiguracji związany jest jedynie z napięciem powierzchniowym na granicach faz.

Źródło: Opracowanie własne



Rysunek 5.8. Reprezentatywny przykład pokazujący ewolucję granic separujących fazy (zwężanie po średniej krzywiźnie). Obrazek pokazuje spiralną konfigurację dominującego wzorca ($\mathcal{P}_1$) i jego negatywu. Na początku symulacji (lewy górny), po $9 \cdot 10^6$ krokach (prawy górny), po $2.07 \cdot 10^7$ krokach (lewy dolny), po $3.73 \cdot 10^7$ krokach (prawy dolny). W czasie ewolucji spirala rozwija się sprawiając wrażenie ruchu wokół środka układu.

Źródło: Opracowanie własne

5.4. Potencjalne możliwości prowadzenia dalszych badań

Badany układ jest w rzeczywistości bardzo ogólny, wiele kwestii związanymi z tego typu systemami stanowi nadal zagadkę. Największym wyzwaniem teoretycznym w tej chwili wydaje się znalezienie opisu dla ruchu punktu współistnienia faz w funkcji temperatury. Wiadomo, że ma na to wpływ struktura korelacji między wzorcami oraz czynnik entropii i że trzeba rozwiązać zagadnienie równości energii swobodnych poszczególnych wzorców (co nie wydaje się być rzeczą prostą). Możliwość znalezienia położenie punktu współistnienia faz może znacznie polepszyć działanie sieci odtwarzających wzorce (być może pozwoliłoby to na skonstruowanie dynamicznego algorytmu uczenia, który dostosowywałby wagi sieci w zależności od temperatury. Ciekawym zagadnieniem są także przewidywane teoretycznie fazy dominowane entropijnie występujące w punkcie krytycznym, w których układy minimalizują energię swobodną poprzez maksymalizację entropii[8]). Badań wymaga także możliwość opisania sieci lokalnie Hopfieldowskich w terminach potencjałów Kaca (dotychczasowe wyniki wskazują, iż może być to dobra droga). Wyniki uzyskane w powyższych zagadnieniach powinny być ostatecznie przeniesione do sieci w pełni Hopfieldowskich, w kontekście modyfikacji wag sieci w zależności od lokalnej struktury korelacji uczonych wzorców (a także do sieci ogólnie rekurencyjnych, rozwiązujących ogólne problemy[9]). Interesującym wydaje się także rozszerzenie tych wyników na sieci Hopfielda ze strukturą typu *small world network* (sieci małego świata), których zdolność odtwarzania wzorców jest obecnie badana [6].

Tylko do celów edukacyjnych

[8] Wymaga to symulacji z większą ilością wzorców i dobrej znajomości położenia punktu współistnienia faz.

[9] Wymaga to przetłumaczenia uzyskanych wyników z języka wzorców, na bardziej ogólny grunt co może być ciekawym wyzwaniem (niekoniecznie możliwym w ogólności).

ROZDZIAŁ 6

Dodatki

6.1. Szczegóły implementacyjne programu

W celu zasymulowania działania opisanego w tej pracy układu stworzony został zestaw programów w języku C++. Do wizualizacji wykorzystany został język Java. Źródła wszystkich używanych programów dołączone są na płytce CD (załącznik do pracy), w tym rozdziale podkreślone zostaną jedynie pewnie z zastosowanych rozwiązań.

- Reprezentacja wag.
  Ilość pamiętanych wag rośnie w przybliżeniu z kwadratem ilości neuronów. W przypadku sieci z lokalną strukturą połączeń wzrost ten jest powolniejszy, mimo wszystko jednak pamiętanie wszystkich wag byłoby sporym marnotrastwem. Okazuje się jednak, że w przypadku rozważanego systemu wystarczyło pamiętać jedynie niewielki procent wag, które ze względu na periodyczny charakter wzorców, również powtarzają się w układzie. Aby dodatkowo zoptymalizować program, wszystkie wagi zostały zapamiętane w jednowymiarowej tablicy, która z powodzeniem mogla się zmieścić do pamięci cache procesora. Procedura wypełniającą tę tablicę przedstawia się następująco:

```
void BoltzmannMachine::initWeights()
{
    for (unsigned int i = 0; i < Pattern_Size; i++)
       for (unsigned int j = 0; j < Pattern_Size; j++)
          for (unsigned int k = 0; k < Pattern_Size; k++)
             for (unsigned int l = 0; l < Pattern_Size; l++) {
                weights[i * Pattern_Size_qub +
                j * Pattern_Size_sqr + k * Pattern_Size + l] = 0;

                for (unsigned int x = 0; x < Patterns.size(); x++)
                   weights[i  * Pattern_Size_qub +
                             j  * Pattern_Size_sqr +
                             k * Pattern_Size + l] +=
                                  (Patterns[x].getValue(i, j) *
                                   Patterns[x].getValue(k, l) * Forces[x]);
             }

}
```

  Następnie skorzystanie z wagi wymaga już jedynie obliczenia odpowiedniego indeksu:

```
double BoltzmannMachine::getWeight(int x, int y, int xx, int yy)
{
        if ((x == xx) && (y == yy))
                return 0;
        else {
```


Tylko do celów edukacyjnych

```
                    int x_ = ((x % Pattern_Size) + Pattern_Size) % Pattern_Size;
                    int y_ = ((y % Pattern_Size) + Pattern_Size) % Pattern_Size;
                    int xx_ = ((xx % Pattern_Size) + Pattern_Size) % Pattern_Size;
                    int yy_ = ((yy % Pattern_Size) + Pattern_Size) % Pattern_Size;

                    return weights[x_ * Pattern_Size_qub + y_ * Pattern_Size_sqr +
                                   xx_ * Pattern_Size + yy_];
            }
    }
```

- Wyliczanie pola lokalnego.
  Głównym i najczęściej wykonywanym fragmentem programu było wyliczenie pola lokalnego dla danego neuronu. Ze względu na wcześniej ustaloną strukturę sąsiedztwa, pętla wyliczająca odpowiednie pole została rozwinięta. Pozwala to uniknąć narzutu związanego z przygotowaniem pętli, który ze względu na jej niewielką długość mógł stanowić istotny procent czasu wykonania programu.

6.2. Płytka CD z filmami zawierającymi ewolucję układu

Do pracy załączona została płytka CD zawierająca dodatkowe dane, obrazki, źródła programów oraz filmy przedstawiające ewolucję rozważanego układu w różnych warunkach.

```
filmy/                          # filmy
  /beta005                       # filmy przy temperaturze odwrotnej 0.005
  /beta03                         # filmy przy temperaturze odwrotnej 0.03
  /beta1                           # filmy przy temperaturze odwrotnej 1.0
  /beta10                         # filmy przy temperaturze odwrotnej 10.0
  /punkt_wielofazowy      # filmy w warunkach współistnienia faz
narzedzia/                   # narzędzia używane w wizualizacji
  /wiz_2d                         # dwuwymiarowej
  /wiz_3d                         # trójwymiarowej
obrazki                       # bitmapy z wykresami diagramów fazowych
praca/                        # elektroniczna wersja tej pracy
  /src                              # źródło TeX tej pracy
program                      # źródła głównego programu symulującego
dane                          # dane uzyskane z symulacji
prezentacje                 # prezentacje na referaty związane z tą pracą
```

Filmy zostały przygotowane w formacie DivX (mpeg4).

6.3. Podziękowania

Autor niniejszej pracy chciałby podziękować twórcom biblioteki "rng" (Random Number Generator), oraz biblioteki wizualizacyjnej VISAD oraz technologii Java3d. Bez ich pracy i możliwości skorzystania z niej za darmo (na warunkach licencji GPL itp.) wyniki zaprezentowane tutaj byłyby zapewne o wiele skromniejsze. Chciałbym też podziękować profesorowi Włodzisławowi Duchowi za zaszczepienie we mnie chęci do badania układów neuronowych,

Tylko do celów edukacyjnych

doktorowi Tomaszowi Schreiberowi za skierowanie moich zainteresowań w kierunku termodynamiki i cierpliwe odpowiedzi na setki moich pytań. Podziękowania należą się także profesorowi Piotrowi Bale, za możliwość skorzystania z klastra obliczeniowego będącego pod jego opieką, a także za sugestie na temat modelowania komputerowego.

Tylko do celów edukacyjnych

Bibliografia

[1] E. Aarts, J. Korst (1989) *Simulated annealing and Boltzmann machines. A stochastic approach to combinatorial optimization and neural computing.* Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Ltd., Chichester.

[2] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski (1985) *A learning algorithm for Boltzmann machines.* Cognitive Science, 9, 147–169.

[3] S. M. Bohte (2003) *Spiking Neural Networks.* PhD thesis, University of Leiden.

[4] B. Borowik (2002) *Pamięci asocjacyjne.* Wydawnictwo MIKOM, Warszawa 2002.

[5] A. Bovier (2001) *Statistical Mechanics of Disordered Systems.* Lecture Notes Weierstrass-Institut fur Angewandte und Stochastik, Mohrenstrasse 39, Berlin.

[6] N. Davey, B. Christianson, R. Adams (2004) *High Capacity Associative Memories and Small World Networks.* IEEE Proceedings IJCNN 2004, Vol 1, 177-182.

[7] S. Diederich, M. Opper (1987) *Learning of Correlated Patterns in Spin-Glass Networks by Local Learning Rules*, Physical Review Letters **58**, no. 9, 949-952.

[8] H.-O. Georgii, *Gibbs Measures and Phase Transitions*, de Gruyter, Berlin, New York, 1988.

[9] W. Duch, J. Korczak (1998) *Optimization and global minimization methods suitable for neural networks* Neural Computing Surveys

[10] W. Duch, J. Korbicz, L. Rutkowski, R. Tadeusiewicz (2000) *Biocybernetyka i inżynieria biomedyczna 2000 - tom 6 Sieci neuronowe.* Akademicka Oficyna Wydawnicza Exit,Warszawa 2000.

[11] R. Ellis (1985) *Entropy, large deviations, and statistical mechanics.* Fundamental Principles of Mathematical Sciences 271, Springer-Verlag, New York 1985.

[12] I. M. Gelfand, G. E. Shilov (1964) *Genralized functions and operations with them.* Academic Press, New York and London.

[13] H. O. Georgii, O. Haeggstroem, C. Maes (2001) The random geometry of equilibrium phases. *Phase transitions and critical phenomena*, **18**, 1-142, Academic Press, San Diego.

[14] O. Haeggstroem (2002) *Finite Markov Chains and Algorithmic Applications*, London Mathematical Society Student Texts, **52**, Cambridge University Press, Cambridge.

[15] S. Haykin (1994) *Neural Networks, a Comprehensive Foundation.* Macmillan College Publishing Company, New York 1994.

[16] D. O. Hebb (1949) *The organization of behavior.* New York, J. Wiley.

[17] G. E. Hinton, T. J. Sejnowski (1983) *Optimal perceptual inference.* Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Washington DC.

Tylko do celów edukacyjnych

[18] G. E. Hinton, T. J. Sejnowski (1986) *Learning and relearning in Boltzmann machines.* In Rumelhart, D. E. and McClelland, J. L., editors, Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations, MIT Press, Cambridge, MA.

[19] J. J. Hopfield (1982) *Neural networks and physical systems with emergent collective computational abilities* Proc Natl Acad Sci U S A. 1982 April; 79(8): 2554-2558.

[20] I. Kanter, H. Sompolinsky, Associative Recall of Memory without Errors, *Phys. Rev. A*,**35**,380-392.

[21] R. A. Kosiński (2002) *Sztuczne sieci neuronowe - dynamika nieliniowa i chaos.* Wydawnictwa Naukowo Techniczne, Warszawa 2002.

[22] J. Mańdziuk Sieci neuronowe typu Hopfielda. Akademicka oficyna wydawnicza EXIT, Warszawa 2000.

[23] W. S. McCulloch, W. H. Pitts (1943) *A logical calculus of the ideas immanent in nervous activity.* Bulletin of Mathematical Biophysics, 5:115-133.

[24] R. A. Minlos, *Introduction to Mathematical Statistical Mechanics*, AMS University Lecture Series **19**, 2000.

[25] D. Parker (1982) *Learning-Logic.* Invention Report S81-64, File 1, Office of Technology Licensing, Stanford University.

[26] R. Peierls (1936) *On Ising model of ferromagnetism.* Proc. Cambridge Phil. Soc 32 (1936, 477-481).

[27] P. Pereto (1992) *An introduction to the Modeling of Neural Networks.* Cambridge University Press 1992.

[28] Personnaz, L., Guyon, I., Dreyfus, G., *J. Phys. (Paris) Lett.* **46**, L359 (1985)

[29] F. Piękniewski, L. Rybicki *Visualizing and Analyzing Multidimensional Output from MLP Networks via Barycentric Projections*, Lecture Notes in Artificial Inteligence vol 3070, 2004.

[30] F. Piękniewski, T. Schreiber *Phase diagrams in locally Hopfield neural networks in presence of correlated patters* (2005, subm. to IJCNN'05).

[31] E. Presutti *From Statistical Mechanics towards Continuum Mechanics*, Max-Planck Institute, Leipzig, 1999.

[32] J. Propp, D. Wilson (1996) *Exact Sampling with Coupled Markov Chains and Applications to Statistical Mechanics.* Random Structures and Algorithms 9, 232-252.

[33] R. Rojas (1996) *Neural Networks: A Systematic Introduction.* Springer-Verlag New York, Inc.

[34] D. E. Rumelhart, G. E. Hinton, R.J. Williams (1986) *Learning internal representations by error propagation, in Parallel Distributed Processing* Explorations in the Microstructures of Cognition, Vol.I, MIT Press, pp. 318-362.

[35] Ya. G. Sinai, *Theory of Phase Transitions: Rigorous Results*, Pergamon Press, 1982.

Tylko do celów edukacyjnych

[36] P. Werbos (1974) *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences.* PhD thesis, Harvard University.

[37] D. Wilson (2000) *How to couple from the past using a read-once source of randomness.* Random Structures and Algorithms 16, 85-113.

[38] M. Zahradnik, An Alternate Version of Pirogov-Sinai Theory, *Communications in Mathematical Physics* **93** (1984) 559-581.

Tylko do celów edukacyjnych

Spis tabel

5.1. Porównanie danych teoretycznych i obliczeniowych 57
5.2. Ruch punktu wielofazowego (rezultaty obliczeń) 57
5.3. Objętości (w %) obszarów czystych faz w diagramach fazowych (rezultat obliczeń). 58

Tylko do celów edukacyjnych

Spis rysunków

1.1. Schemat Neuronu formalnego 9

2.1. Schemat pięcioelementowej sieci rekurencyjnej 13
2.2. Przykład funkcji (energetycznej) zadanej na odcinku $[0, 1]$ (czerwony wykres najniżej) oraz seria gęstości prawdopodobieństwa miar Gibbsa dla coraz większych parametrów β. Łatwo zaobserwować, że miary te coraz bardziej skupiają się na minimum funkcji energetycznej. 23

3.1. Przykładowy łańcuch Markowa zawierający trzy stany przechodnie (niebieskie) dwa stany absorbujące (czerwone) i trzy stany tworzące klasę rekurencyjną (zielone). Etykiety na krawędziach oznaczają prawdopodobieństwa przejść ze stanu w którym krawędź się rozpoczyna do stanu gdzie krawędź się kończy. . 27
3.2. Schematyczny przebieg algorytmu Proppa-Wilsona na przestrzeni fazowej o 5 stanach. W kolejnych krokach równolegle liczone drogi ewolucji się sklejają, by ostatecznie zakończyć się w jednym stanie. Stan ten zgodnie z istotą algorytmu jest wylosowany z rozkładu stacjonarnego. 33

4.1. Przykładowy wygląd fragmentu konfiguracji konturowej. 45
4.2. Diagram fazowy dla modelu Isinga 47

5.1. Przykłady skorelowanych wzorców użytych w symulacji. Najbardziej lewy ($\mathcal{P}_1$) jest wzorcem silnie dominującym (na dalszych diagramach fazowych będzie oznaczany kolorem czerwonym). Drugi w kolejności z lewej ($\mathcal{P}_2$, oznaczany kolorem zielonym na diagramach) jest silnie skorelowany z pierwszym, razem tworzą "koalicję" wspomagając się przeciw trzeciemu wzorcowi ($\mathcal{P}_3$, oznaczany niebieskim kolorem na diagramach), który ma najmniejszą korelację z innymi. Czwarty wzorzec ($\mathcal{P}_4$) ma również niewielką korelację z innymi. 51
5.2. Typowa sytuacja występująca w czasie symulacji. Jeden silny wzorzec ($\mathcal{P}_1$) oraz plama jego negatywu w środku. Warto zauważyć, że algorytm zliczający pozycje niezgodności w komórkach podstawowych, zliczy jedynie pozycje w komórkach zawierających zarówno wzorzec jak i negatyw (których jest relatywnie bardzo mało), co skutkuje tym, że parametr D_{glob} będzie bardzo niewielki. . . 52
5.3. Przykład diagramu fazowego dla czterech wzorców ($\mathcal{P}_1$,$\mathcal{P}_2$,$\mathcal{P}_3$,$\mathcal{P}_4$,$\beta = 0.5$). Przedni fragment wykresu jest przezroczysty dla lepszego wyeksponowania punktu współistnienia faz. Obszary stabilności pojedynczych wzorców są dość regularne, a ich granice dość ostre. 54
5.4. Przykład diagramu fazowego dla trzech wzorców ($\mathcal{P}_1$,$\mathcal{P}_2$,$\mathcal{P}_3$). Każde pole koloru oznacza obszar gdzie dokładnie jeden wzorzec jest stabilny. Miejsce w którym wszystkie pola się stykają jest punktem współistnienia faz. 55
5.5. Przykładowe cięcia przez diagram fazowy (cztery wzorce $\mathcal{P}_1$,$\mathcal{P}_2$,$\mathcal{P}_3$,$\mathcal{P}_4$ z rysunku 5.1, $\beta = 0.5$). Widać na nich wyraźnie położenie punktu współistnienia faz, geometrię płaszczyzn współistnienia dwóch faz i krzywych współistnienia trzech faz. Nieregularności granic międzyfazowych to pozostałość po próbkowaniu Monte-Carlo i niedoskonałości tej metody. Generalnie im więcej próbek zebrano tym gładsze były granice. 55

Tylko do celów testowych

5.6. Ewolucja sieci w warunkach ścisłej dominacji jednego ze wzorców ($\mathcal{P}_1$) w średniej temperaturze ($\beta = 0.2$), w początkowej fazie symulacji (po lewej) oraz po $9 \cdot 10^4$ krokach. Elipsa i kółko utworzone z niestabilnych wzorców ($\mathcal{P}_3$,$\mathcal{P}_4$) są gwałtownie pochłaniane przez stabilny wzorzec. Proces ten jest szybki i gwałtowny, w przeciwieństwie do powolnych procesów brzegowych pokazanych na rysunku 5.7. 59
5.7. Stan sieci w warunkach współistnienia faz w niskiej temperaturze ($\beta = 1$), stan początkowy (na lewo) i po $3.39 \cdot 10^7$ krokach (na prawo). Elipsa i kółko utworzone ze wzorców zwężają się według średniej krzywizny. Proces ten jest powolny i spokojny, gdyż naddatek energii konfiguracji związany jest jedynie z napięciem powierzchniowym na granicach faz. 59
5.8. Reprezentatywny przykład pokazujący ewolucję granic separujących fazy (zwężanie po średniej krzywiźnie). Obrazek pokazuje spiralną konfigurację dominującego wzorca ($\mathcal{P}_1$) i jego negatywu. Na początku symulacji (lewy górny), po $9 \cdot 10^6$ krokach (prawy górny), po $2.07 \cdot 10^7$ krokach (lewy dolny), po $3.73 \cdot 10^7$ krokach (prawy dolny). W czasie ewolucji spirala rozwija się sprawiając wrażenie ruchu wokół środka układu. 60

Tylko do celów edukacyjnych

Skorowidz

Algorytm
— Minimalizacji globalnej, 11
— multistart, 12
— P*, 12
— Proppa-Wilsona, 33
— Random walk, 11
— Simulated Annealing, 12
— Tabu search, 11
Atraktor, 7
— pasożytniczy, 15, 19
Autoasocjator, 15, 51

Czas
— refrakcji, 13

Diagram fazowy, 48, 55
Dynamika
— asynchroniczna, 14
— Glaubera, 14
— hybrydowa, 14
— Little'a, 14
— synchroniczna, 14

Energia
— konfiguracji, 14
— ograniczeń, 19
— optymalizacji, 19
— specyficzna, 49, 57

Faza
— czysta, 45
— mieszana, 45
Fizyka
— statystyczna, 36

Gaz
— klasyczny, 37
— Konfiguracyjny, 38

Katastrofa zapominania, 19
Klasa rekurencyjna, 27
Komórka podstawowa, 51
Kontur, 45
Kryterium Dobrushina, 45

Markowa
— łańcuch, 8, 26
— aperiodyczny łańcuch, 30
— nieprzywiedlny łańcuch, 30
— odwracalny łańcuch, 29
— proces, 26
Maszyna Boltzmanna, 22
Miara
— DLR, 42
— ekstremalna DLR, 44
— Gelfanda-Leraya, 39
— Gibbsa, 41
— graniczna DLR, 42
— graniczna Gibbsa, 42
— kanoniczna, 40
— mikrokanoniczna, 40

Perceptron, 10
Perturbacja Hamiltonianu, 49
Pole
— lokalne, 14
— wypadkowe, 14
— zewnętrzne, 13
Potencjał
— Lenarda-Jonesa, 37
Problem
— kolorowania grafu, 19
— komiwojażera, 21
— NP-zupełny, 19
— planu zajęć, 20
Proces
— stochastyczny, 26
Przejście fazowe, 45, 47
Przestrzeń fazowa, 7
Punkt współistnienia faz, 49, 57

Równanie
— dokładnego bilansu, 29
Rozkład
— mikrokanoniczny, 40
Rozkład stacjonarny, 28

Sieć
— Hopfielda, 13, 15
— lokalnie Hopfieldowska, 51
— MLP, 11
— Rekurencyjna, 13
Stan bazowy, 49

Tylko do celów edukacyjnych

Temperatura, 23
— Curie, 39
— odwrotna, 23
Teoria Pirogova-Sinaia, 49
Trajektoria, 7

Układ dynamiczny, 7

Własność
— zapominania, 26

Zagadnienie obliczeniowe, 6
Zasada
— wariacyjna, 41
— zachowania energii, 38
— zachowania objętości przestrzeni fazowej, 38
Zjawiska
— objętościowe, 59
— powierzchniowe, 59

Tylko do celów edukacyjnych