



Departement Natuurkunde

Leerstrategieën voor het Binaire Perceptron

Proefschrift voorgelegd tot het behalen van de graad van
doctor in de wetenschappen groep natuurkunde
aan het Limburgs Universitair Centrum te verdedigen door

BART VAN ROMPAEY

Promotor: Prof. dr. M. Bouten

1999

D/1999/2451/10

536.9

UNIVERSITEITSBIBLIOTHEEK LUC



03 04 00521503

990520



17 MEI 1999

536.9

VANR

1999

.luc.luc.luc.



Departement Natuurkunde

Leerstrategieën voor het Binaire Perceptron

990520

Proefschrift voorgelegd tot het behalen van de graad van
doctor in de wetenschappen groep natuurkunde
aan het Limburgs Universitair Centrum te verdedigen door



17 MEI 1999

BART VAN ROMPAEY

Promotor: Prof. dr. M. Bouten

1999

D/1999/2451/10

Aan mijn grootvader E. Van Rompaey[†]

Dankwoord

In de eerste plaats dank ik mijn promotor Marc Bouten voor zijn uitstekende en geduldige begeleiding en voor de vele goede raad die hij mij gegeven heeft. Verder wens ik al de andere leden van de groep Theoretische Natuurkunde (Roger Serneels, Carlo Vanderzande, Chris Van den Broeck, Frank Daerden, Mauro Copelli, Ioana Bena-Birzu, Peter Leoni en Jef Hooyberghs) te bedanken voor hun bijdragen en de aangename werksfeer. Een speciale vermelding verdienen Mauro Copelli, hij is een zeer aangename "bureau-genoot" zowel binnen als buiten het LUC, alsook Geert Jan Bex om menig computerprobleem uit de weg te ruimen. Verder wens ik ook twee buitenlandse gasten, Lueder Reimers en Doerthe Malzahn, te bedanken voor hun bijdragen en aangename discussies.

Tevens zijn er nog een heleboel mensen die ik wens te bedanken voor zaken die niet rechtstreeks gerelateerd waren aan mijn onderzoek. Herman Janssen dank ik voor de gelegenheid die ik heb gekregen om één van zijn cursussen te mogen begeleiden. Dat was een zeer aangename en leerrijke ervaring die ik iedereen zou toewensen. Verder gaat mijn dank uit naar vele anderen (i.h.b. M. Van Gastel, L. Vandeurzen, K. Meykens, J. Van Olmen, L. Bruckers, G. Claeskens, P. De Grootte). Zij hebben allen bijgedragen tot een aangename leefomgeving in en buiten het LUC en elk van hen weet wel hoe en waarom. Ook gaat mijn oprechte dank uit naar de dames van het secreteriat WNI, (H. Wijnen, C. Wijnants, V. Mebis, M. Machiels en A. Hermans) voor alle hulp met vele administratieve problemen.

Tenslotte zijn er nog een aantal mensen die ik wil bedanken maar die niet verbonden zijn met het LUC. In de eerste plaats zijn dat natuurlijk mijn ouders, J. Van Rompaey en J. Claes. Zonder hen had ik nooit zo ver kunnen komen. Verder is er nog mijn grootvader E. Van Rompaey[†] zonder dewelke ik nooit natuurkundige was geworden. And last but not least zijn er nog Joris Claes, Els Staes, Beate Luyckx, Nadia Berden, en Géisa Fernandes. Zij weten ook wel waarom.

Het F.W.O. Vlaanderen wens ik te danken voor financiële steun.

Inhoudsopgave

1	Probleemstelling	1
1.1	Inleiding	1
1.2	Het associatief geheugenprobleem	2
1.2.1	Hopfields opvatting	2
1.2.2	Het model	3
1.2.3	Gekende resultaten	7
1.3	Overzicht	10
2	Het precursor idee	13
2.1	Inleiding	13
2.2	Kwaliteitsmaat voor een continue precursor	14
2.3	De gemeenschappelijke distributie $P(J, B)$	15
2.3.1	De keuze van de leerregels	18
2.4	Resultaten voor $P(j, b)$	18
2.4.1	De zadelpuntvergelijkingen voor het precursor systeem	19
2.4.2	De zadelpuntvergelijkingen voor het MSB systeem	20
2.4.3	De zadelpuntvergelijkingen voor r en $\hat{r}$	21
2.4.4	De uitdrukking voor $P(j, b)$	22
3	Leren op de hypersfeer	27
3.1	Inleiding	27
3.2	Het $V(\lambda)$ /MSB systeem	28
3.3	De MSN als precursor	31
3.4	De Optimale Precursor	32
3.4.1	Een quasi-optimale potentiaal	35
3.5	Simulaties	36

INHOUDSOPGAVE

4	Leren in de hyperkubus	41
4.1	Inleiding	41
4.2	Het $V(\lambda)$ /MSB systeem	43
4.3	De MSC als Precursor	47
4.4	De optimale precursor	50
4.4.1	Een quasi-optimale potentiaal	51
4.5	Simulaties	57
4.5.1	Oplossing voor de ± 1 ontarding.	58
4.5.2	Resultaten	59
5	Optimale leerstrategie voor het binair perceptron	63
5.1	Inleiding	63
5.2	De methode van branch en bound	64
5.2.1	Hoe efficient programmeren?	66
5.3	De leerstrategie	68
5.4	Simulaties	70
6	Conclusie	75
A	Replica symmetrie berekening voor $P(j, b)$	77
A.1	Introductie van replicas	77
A.2	De RS-ansatz	80
A.3	De limiet $n \rightarrow 0$	82
A.4	De zadelpuntvergelijkingen voor het precursor systeem	85
A.5	De zadelpuntvergelijkingen voor het MSB systeem	86
A.6	De zadelpuntvergelijkingen voor r en $\hat{r}$	88
A.7	De vorm van $P(j, b)$	90
A.8	Wat onder globale constraints?	91
B	$V(\lambda)$/MSB op de hypersfeer	93
B.1	De zadelpuntvergelijkingen voor $V(\lambda)$	93
B.2	De zadelpuntvergelijkingen voor r en $\hat{r}$	95
B.3	Berekening van $P(j, b)$ en de kwaliteitsmaat f_x	97
C	$V(\lambda)$/MSB in de hyperkubus	101
C.1	De zadelpuntvergelijkingen voor $V(\lambda)$	101
C.2	De zadelpuntvergelijkingen voor r en $\hat{r}$	103
C.2.1	Betekenis van de zadelpuntvergelijking voor r	104

INHOUDSOPGAVE

C.3	Berekening van $P(j, b)$ en de kwaliteitsmaat f_x	105
C.4	Berekening van $P(j)$ in de hyperkubus	109
C.5	De kostfunctie $E(\mathbf{J})$ met de perturbatieterm $\rho \mathbf{J}^2$	109
D	Over de convexiteit van kostfuncties	113
D.1	Definitie van convexiteit van een functie	113
D.2	De kostfunctie bekomen door de potentiaal $V_{qo}(\lambda)$ op de hypersfeer	113
D.2.1	De kostfunctie bekomen door de potentiaal $V_{qo}(\lambda)$ in de hyperkubus	115
E	Formularium	117
E.1	Notaties	117
E.2	Enkele eenvoudige functies	118
E.3	De Fouriertransformatie	118
E.4	Enkele identiteiten.	118

INHOUDSOPGAVE

Lijst van figuren

1.1	Volledig geconnecteerd netwerk	4
1.2	Het perceptron	7
1.3	De geheugencapaciteit $\alpha_c(\kappa)$	8
2.1	Oplossing van de MSB vergelijkingen	22
2.2	De entropie voor het binair systeem	23
3.1	$f_1(\alpha)$ op de hypersfeer	32
3.2	$f_x(\alpha)$ op de hypersfeer	33
3.3	De functie $h(r)$	34
3.4	De quasi-optimale potentiaal op de hypersfeer	36
3.5	f_1 en f_x op de hypersfeer met de quasi optimale potentiaal . . .	37
3.6	Simulaties op de hypersfeer voor $N = 50$	39
4.1	Hypersfeer vs hyperkubus	42
4.2	Oplossing van de MSC vergelijkingen	48
4.3	Vergelijking van de MSN, de MSC en de MSB	49
4.4	$f_1(\alpha)$ voor de MSC precursor	50
4.5	$f_x(\alpha)$ voor de MSC precursor	51
4.6	Vergelijking van MSC en V_{qo} op de hypersfeer	52
4.7	Oplossing van de vergelijkingen voor V_{qo} in de hyperkubus . . .	53
4.8	$f_1(\alpha)$ voor de V_{qo} -precursor in de hyperkubus	54
4.9	Simulaties voor volledig knippen van precursors	55
4.10	$f_x(\alpha)$ voor de V_{qo} -precursor in de hyperkubus	56
4.11	Simulaties voor $N = 40$	60
4.12	Simulaties voor $N = 75$ en $N = 100$	61
5.1	Voorstelling van een binaire boom	65
5.2	$f_x(\alpha)$ met $x = 0.2$ voor V_{qo} in de hyperkubus	69

LIJST VAN FIGUREN

5.3	Branch en bound voor $N=40$	71
5.4	Branch en bound voor $N=50$	73
C.1	Nieuw integratiedomein	108
D.1	Convexiteit op de hypersfeer	114

Hoofdstuk 1

Probleemstelling

1.1 Inleiding

De ontwikkeling van computers heeft onze wereld totaal veranderd. Vandaag de dag is het ondenkbaar zich een wereld zonder deze merkwaardige machines in te beelden. Sedert hun ontstaan verlichten ze ons werk: computers voeren snel en betrouwbaar berekeningen uit en zijn uitmuntend in het uitvoeren van repetitieve taken. Maar toch is het zo dat er nog vele taken zijn waar de hersenen veruit superieur zijn. Een mooi voorbeeld is het verwerken van visuele informatie: ons vermogen om voorwerpen of gezichten te herkennen kent geen vergelijk, zelfs niet met de snelste supercomputers. Merkwaardig genoeg zijn de bouwstenen van de hersenen — de neuronen — veel minder performant dan hun tegenhangers in de computerarchitectuur (de transistor). De hersenen zijn verder heel robuust en fout tolerant, het leert zonder expliciet voorgeprogrammeerd te moeten worden en kan informatie verwerken die onvolledig of zelfs inconsistent is.

Het lijkt er steeds meer en meer op dat de toename van de relatieve grootte van de hersenen van een zoogdier een beslissende factor is geweest in het evolutieproces dat geleid heeft tot de bovenvermelde eigenschappen [33]. Gebruik makend (of misbruik makend) van metaforen uit de statistische fysica: het lijkt erop dat de grootte van de hersenen een van de macroscopische parameters is geweest die het begin van nieuw gedrag controleert, zoals in een fase overgang. Een groot aantal onafhankelijke kleine CPU eenheden heeft blijkbaar voordelen. Elk van deze eenheden is bovendien geconnecteerd met een groot aantal andere eenheden. Elk van hen zendt en ontvangt informatie. Het

1. Probleemstelling

blijkt dus dat de natuur tijdens het evolutieproces een eenvoudige oplossing gevonden heeft om met zeer trage en oude "chips" een krachtige "computer" te bouwen. Het volstaat een groot aantal van hen te nemen (ongeveer 10^{11} in de hersenen van een mens) en ze te connecteren met een groot aantal andere op een min of meer willekeurige manier. Een cruciaal ingrediënt is wel dat de hersenen in staat zijn de connecties aan te passen tijdens een leerproces. Dit resulteert in de eigenschap dat de hersenen zich kunnen aanpassen aan een nieuwe omgeving.

Het behoeft geen verder betoog dat het enorme voordelen zou bieden om een dergelijke computer te bouwen, gebaseerd op de structuur en werkwijze van de dierlijke hersenen. Een groot deel van het beginnend onderzoek in artificiële intelligentie en neurale netwerken in het bijzonder is gewijd aan patroon herkenning [12, 1, 10]. Dit fenomeen wordt gezien als een van de hoekstenen van intelligent gedrag. Wij zijn heel gemakkelijk in staat van een persoon die we in jaren niet meer hebben gezien terug te herkennen, zelfs als deze persoon wezenlijk veranderd is. Het is voor ons tevens kinderspel om de volgende zin aan te vullen: "To be or..." of om $E = mc^3$ te corrigeren. Het is pas vrij recent dat het probleem van patroon herkenning met behulp van neurale netwerken de interesse heeft opgewekt bij de statistische fysici. Eén van de papers die daartoe bijgedragen heeft is het baanbrekend werk van J. J. Hopfield [12]. In de nu volgende paragrafen zal ik dieper ingaan op dergelijke associatieve geheugens. Ik zal een model invoeren van een dergelijk geheugen en hiermee tevens het probleem schetsen dat ik in deze thesis wens te behandelen.

1.2 Het associatief geheugenprobleem

1.2.1 Hopfields opvatting

Hopfield zag in dat een associatief geheugen verscholen zit in het natuurlijk gedrag van vele fysische systemen. Beschouw de tijdsevolutie van een systeem dat kan beschreven worden met een set van algemene coördinaten waarbij een punt in die coördinatenruimte dan de toestand van het systeem voorstelt op een bepaald tijdstip. De bewegingsvergelijkingen van dit systeem beschrijven dan een "baan" in de toestandsruimte. Tijdens de evolutie van het systeem zijn verscheidene soorten "banen" mogelijk maar de systemen die kunnen gebruikt worden als associatieve geheugens zijn diegene die evolueren naar lokaal stabiele toestanden vanuit gelijk welke toestand in hun nabije omgeving.

1.2. Het associatief geheugenprobleem

Neem als voorbeeld een fysisch systeem dat kan beschreven worden door N coördinaten $S_1, S_2, \dots, S_N$ welke de componenten zijn van de toestandsvektor $\mathbf{S}$. Veronderstel verder dat dit systeem p toestanden heeft die lokaal stabiel zijn, zeg $\xi^1, \xi^2, \dots, \xi^p$. Dus als we het systeem een beginvoorwaarde $\mathbf{S}$ geven die zich voldoende dicht bij een van deze p toestanden bevindt, $\mathbf{S} = \xi^\mu + \Delta$, dan zal het systeem in de tijd evolueren tot $\mathbf{S} \approx \xi^\mu$. In deze beschrijving komen de p stabiele toestanden overeen met de informatie die we wensen op te slaan in het systeem. Het startpunt $\mathbf{S} = \xi^\mu + \Delta$ symboliseert de gedeeltelijke kennis van het objekt ξ^μ en de dynamica van het systeem produceert dan de gewenste totale informatie ξ^μ .

Gelijk welk fysisch systeem waarvan de dynamica gedomineerd wordt door een aanzienlijk aantal lokaal stabiele toestanden kan dus beschouwd worden als een associatief geheugen. Het is bovendien een potentieel nuttig geheugen als de dynamica voldoende "flexibel" is (door het veranderen van parameters) om gelijk welk punt in de toestandsruimte als één van de stabiele toestanden te doen optreden en het aantal stabiele toestanden gevarieerd kan worden. In de volgende paragrafen zal duidelijk worden dat modellen van neurale netwerken uiterst geschikt zijn om dergelijke associatieve geheugens te construeren.

1.2.2 Het model

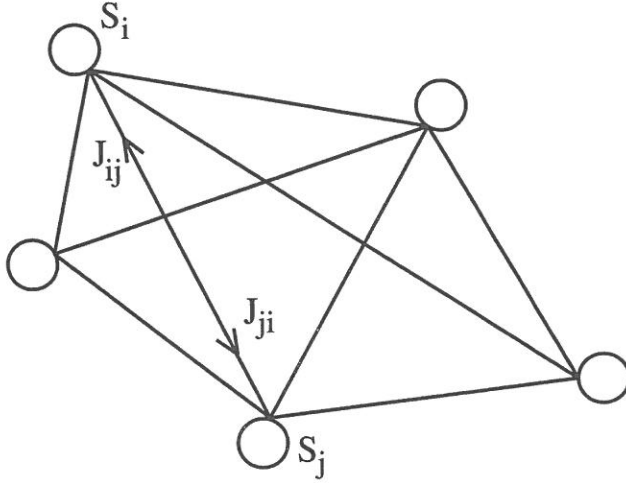
Beschouw een neuraal netwerk dat bestaat uit N neuronen. Elk neuron i kan zich in twee toestanden bevinden $S_i = +1$ en $S_i = -1$. De toestandsvektor $\mathbf{S}$ van dit netwerk is dus een string van N bits. In het algemeen hebben alle neuronen verbindingen met een aantal andere neuronen in het netwerk. De sterkte van zulk een verbinding of synaps tussen twee neuronen i en j noteren we als J_{ij} . Indien neuron i en j niet geconnecteerd zijn is $J_{ij} = 0$. In dit model hebben we een volledig geconnecteerd netwerk waarin elk neuron met alle andere neuronen verbonden is behalve met zichzelf ($J_{ii} = 0$). (Zie figuur 1.1.)

De dynamica van het netwerk waarmee de N -dimensionale toestandsvektor $\mathbf{S}$ verandert is gegeven door:

$$S_i = \text{sign}\left(\sum_j J_{ij} S_j\right). \quad (1.1)$$

De sign-functie is gedefinieerd in (E.4) op pagina 118. Deze regel kan op willekeurige tijdstippen zowel toegepast worden op elk neuron afzonderlijk als

1. Probleemstelling



Figuur 1.1: Voorbeeld van een volledig geconnecteerd netwerk met $N = 5$. De synapsen J_{ij} en J_{ji} tussen neuronen i en j zijn grafisch weergegeven in 1 lijn om de figuur niet te overladen.

op alle neuronen tegelijk. Men spreekt respectievelijk van asynchrone en parallele updates. De connectie J_{ij} kan zowel positief als negatief zijn, het wordt respectievelijk een exciterende en inhiberende synaps genoemd.

Onder de dynamica (1.1) willen we nu $p = \alpha N$ lokaal stabiele toestanden ξ^μ , $\mu = 1 \dots p$ creëren. De parameter α — de geheugenbezetting — is een maat voor het aantal verschillende patronen ξ dat we wensen te memoriseren in het netwerk. Volgens (1.1) zijn deze p toestanden stabiel als de synapsen J_{ij} waarden aannemen zodat:

$$\xi_i^\mu = \text{sign}\left(\sum_j J_{ij} \xi_j^\mu\right) \quad \forall i \forall \mu. \quad (1.2)$$

Deze voorwaarden eisen dat ξ_i^μ hetzelfde teken heeft als $\sum_j J_{ij} \xi_j^\mu$, m. a. w. het produkt van beiden moet groter of gelijk zijn aan nul:

$$\sum_j J_{ij} \xi_j^\mu \xi_i^\mu \geq 0 \quad \forall i \forall \mu. \quad (1.3)$$

Dit is een set van Np vergelijkingen die moeten opgelost worden naar de synapsen J_{ij} . Er zijn verschillende algoritmen ontwikkeld om een synapsmatrix

1.2. Het associatief geheugenprobleem

$\{J_{ij}\}$ te vinden, zie o. a. [12, 15, 6]. Het is onmiddellijk duidelijk dat indien men steeds meer en meer patronen wenst te memoriseren met een bepaald algoritme, er een bepaalde waarde van α zal zijn waarop de vergelijkingen (1.3) geen oplossing meer hebben. De waarde van α waarop dit gebeurt wordt de geheugencapaciteit α_c genoemd. Het is het maximaal aantal patronen dat men kan opslaan met een bepaald leeralgoritme. Zo geldt bijvoorbeeld voor de regel van Hebb [1] met random gekozen patronen ξ dat $\alpha_c = 0.14$ en zo kan de pseudo-invers regel [15] N lineair onafhankelijke patronen opslaan ($\alpha_c = 1$). Er zijn echter ook berekeningen uitgevoerd om de maximale waarde van α_c te vinden onafhankelijk van het gebruikte leeralgoritme, zie o. a. [9, 10]. Ik kom hierover verder in dit hoofdstuk nog uitgebreid op terug.

De set vergelijkingen (1.3) garandeert enkel dat de p toestanden stabiel zijn onder de dynamica (1.1). Over de geschiktheid van de gevonden synapsen om één van de p patronen te "herinneren" aan de hand van onvolledige of onjuiste informatie, wordt geen uitspraak gedaan en vraagt een afzonderlijk onderzoek. Het blijkt dat de grootte van de attractiegebieden verwaarloosbaar klein wordt als $\alpha \rightarrow \alpha_c$ [8, 16, 25]. We willen echter dat zelfs nabij saturatie ($\alpha \approx \alpha_c$) er een eindig groot attractiegebied bestaat omheen de gememoriseerde patronen.

Om dit probleem op te lossen zijn er methodes ontwikkeld om de synapsen J_{ij} iteratief aan te passen. Als een gegeven toestand niet convergeert naar het gewenste patroon zal men proberen de connecties te wijzigen zodat deze convergentie verzekerd is [29]. Indien men op een dergelijke manier een attractiegebied wil "graven" in de toestandruimte zal men echter een groot deel van de toestanden van dat attractiegebied moeten onderzoeken op convergentie. Een dergelijke methode is zeer tijdrovend vermits het aantal toestanden die in k bits verschillen van het gememoriseerde patroon toeneemt zoals $N^k/k!$.

Een andere "ruwere" methode om de attractiegebieden te vergroten is vergelijking (1.3) aanpassen tot [19, 8, 16, 21]:

$$\lambda_i^\mu \equiv \frac{1}{\sqrt{N}} \sum_j J_{ij} \xi_j^\mu \xi_i^\mu \geq \kappa > 0 \quad \forall i \forall \mu, \quad (1.4)$$

$$\sum_j J_{ij}^2 = N. \quad (1.5)$$

Het symbool λ_i^μ is hier ingevoerd en wordt de stabiliteit van patroon ξ^μ op site i genoemd. De belangrijkste verandering t. o. v de vergelijking (1.3) is de invoering van de parameter $\kappa > 0$. Deze parameter zorgt voor eindig grote attractiegebieden, ook in de buurt van saturatie. De normeringsvoorwaarde

1. Probleemstelling

(1.5) is noodzakelijk. Zodra we een matrix $\{J_{ij}\}$ kennen die een oplossing is van (1.3) volgt onmiddellijk een oplossing van (1.4) door de matrix $\{J_{ij}\}$ te vermenigvuldigen met een voldoende grote en positieve parameter γ : $\{\gamma J_{ij}\}$. Dergelijke oplossingen zijn echter niet essentieel verschillend. Met de invoering van (1.5) wordt deze triviale ontaarding uitgesloten. Verder is er nog een factor $\frac{1}{\sqrt{N}}$ ingevoerd in (1.4). Deze zorgt ervoor dat de parameter κ van $\mathcal{O}(1)$ is. Immers met de normeringsvoorwaarde (1.5) is $\sum_j J_{ij} \xi_j^\mu \xi_i^\mu$ van $\mathcal{O}(\sqrt{N})$. De bedoeling is nu van een matrix $\{J_{ij}\}$ te vinden die de vergelijkingen (1.4) en (1.5) oplost met een waarde van de parameter κ die zo groot mogelijk is. Hierbij worden dan de attractiegebieden omheen de gememoriseerde patronen zo groot mogelijk gemaakt binnen het formalisme van (1.4) en (1.5). Men zegt dat men de stabiliteiten λ_i^μ van de p patronen wil maximaliseren. De parameter κ is dan de kleinste waarde van deze p stabiliteiten.

In de vergelijkingen (1.4) en (1.5) zijn de verschillende rijen in de matrix $\{J_{ij}\}$ niet gekoppeld als we aannemen dat de matrix niet symmetrisch is. We kunnen dan ook het probleem rij per rij oplossen. Voor een willekeurige rij i kunnen we dan schrijven:

$$\lambda_\mu \equiv \frac{1}{\sqrt{N}} \sum_j J_{ij} \eta_j^\mu = \frac{1}{\sqrt{N}} \mathbf{J} \cdot \boldsymbol{\eta}^\mu \geq \kappa > 0 \quad \forall \mu, \quad (1.6)$$

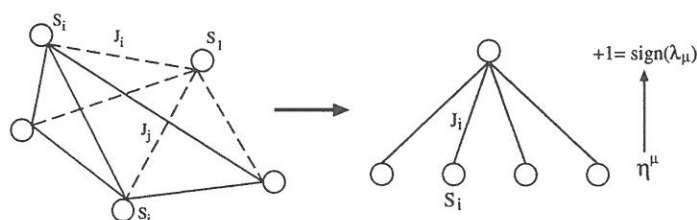
$$\sum_j J_j^2 = \mathbf{J}^2 = N, \quad (1.7)$$

waarbij we $\xi_j^\mu \xi_i^\mu$ genoteerd hebben als η_j^μ . De aandacht kan dus gevestigd worden op de connecties die in één enkel neuron toekomen (i.e. één enkel output neuron geconnecteerd met $N - 1$ input neuronen.) Het volledig geconnecteerd netwerk is nu herleidt tot een perceptron (zie figuur 1.2).

De patronen $\boldsymbol{\eta}^\mu$ worden gememoriseerd met output +1 (wegens (1.6)) en de vergelijkingen zoeken naar het perceptron met maximale stabiliteit. Zoals het probleem nu in de "perceptron vorm" geformuleerd is kunnen we de patronen $\boldsymbol{\eta}^\mu$ zowel binaire als continue waarden laten aannemen. We zullen in al wat volgt aannemen dat de patronen $\boldsymbol{\eta}^\mu$ onafhankelijk identiek gedistribueerde random variabelen zijn met

$$\langle \eta_i^\mu \rangle = 0 \quad \langle \eta_i^\mu \eta_j^\mu \rangle = \delta_{ij}. \quad (1.8)$$

1.2. Het associatief geheugenprobleem



Figuur 1.2: De connecties aangeduid met stippellijn zijn de synapsen die deel uitmaken van het geselecteerde perceptron. De inputpatronen η^μ worden door het perceptron allen geklasseerd als $+1$ indien aan vergelijking (1.6) voldaan is.

1.2.3 Gekende resultaten

Er zijn twee belangrijke vragen die men zich kan stellen bij de vergelijkingen (1.6) en (1.7). Allereerst kan men vragen naar de grootst mogelijke waarde van κ , bij een gegeven waarde van de geheugenbezetting α , waarvoor de vergelijkingen nog oplossingen hebben: $\kappa_{max}(\alpha)$. Of omgekeerd — bij gegeven κ — naar de grootst mogelijke waarde van α : de geheugencapaciteit $\alpha_c(\kappa)$.

Tevens kan men zich afvragen of er een algoritme bestaat dat voor gegeven α het perceptron met maximale stabiliteit als resultaat heeft.

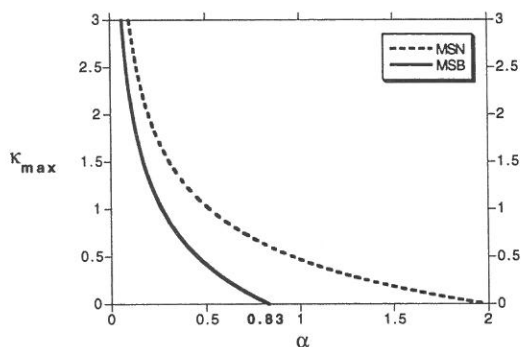
Laten we deze twee vragen even naderbij bekijken. Ik zal bij de bespreking van deze twee vragen onderscheid maken tussen een perceptron met continue koppelingen $\mathbf{J}$, $J_i \in \mathbb{R}$ en een perceptron met binaire koppelingen $\mathbf{B}$, $B_i = \pm 1$.

Continue Koppelingen

De vraag naar de maximale waarde van κ bij een gegeven waarde van α is opgelost door Gardner [9, 10]. In figuur 1.3 is $\kappa_{max} = \kappa_{max}(\alpha)$ in stippellijn weergegeven.

Men ziet dat bij een toenemende geheugenbezetting de grootst mogelijke minimale stabiliteit van de patronen afneemt totdat ze bij $\alpha_c = 2$ nul wordt. Hieruit volgt onmiddellijk dat $\alpha_c = 2$ de maximale waarde is van het aantal patronen dat men kan memoriseren in een netwerk met continue koppelingen. We zullen deze waarde noteren als α_c^{max} . Indien men meer random patronen wenst op te slaan in het netwerk zullen deze niet allemaal een positieve stabiliteit hebben. (κ is dan negatief.)

1. Probleemstelling



Figuur 1.3: De maximale waarde van κ als een functie van α voor zowel continue (stippellijn) als binaire (volle lijn) koppelingen. Indien men de y -as als variabele parameter neemt leest men op de x -as α_c af. ($\alpha_c = \alpha_c(\kappa)$)

Indien $\alpha < \alpha_c^{max} = 2$ bestaat er dus een oplossing voor het maximaal stabiel perceptron bepaald door de vergelijkingen (1.6) en (1.7) en deze oplossing is uniek [9, 10]. Er zijn tevens efficiënte algoritmen gevonden die deze continue koppelvingsvektor $\mathbf{J}$ als resultaat heeft, zoals het AdaTron algoritme [2] en het MinOver algoritme [19]. Deze koppelvingsvektor zal ik voortaan in deze thesis aanduiden met MSN¹.

Binaire koppelingen

De maximale waarde van κ bij een gegeven waarde van α werd berekend door Krauth en Mézard [20]. De functie $\kappa_{max} = \kappa_{max}(\alpha)$ voor binaire koppelingen is in volle lijn weergegeven in figuur 1.3. Men ziet een soortgelijke trend als bij het continue geval, de waarde van de maximaal toegelaten κ daalt bij toenemende waarden van α totdat hij nul wordt bij $\alpha_c^{max} = 0.83$. Dit is opnieuw het maximaal aantal random patronen dat men in een netwerk met binaire koppelingen kan opslaan.

Indien $\alpha < \alpha_c^{max} = 0.83$ bestaat er dus een oplossing voor het maximaal

¹MSN staat voor Maximaal Stabiël Netwerk.

1.2. Het associatief geheugenprobleem

stabiel binair perceptron. Let wel, voor $N \rightarrow \infty$ is deze oplossing niet uniek [20]. Ik zal een dergelijke oplossing in deze thesis voortaan aanduiden met MSB².

Het enige algoritme dat gekend is om het probleem op te lossen is de volledige enumeratie van alle 2^N mogelijke combinaties van koppelingen. Dit is geen polynomiaal algoritme: de computer tijd nodig om een oplossing te vinden neemt exponentieel toe met het aantal koppelingen. Een volledige exploratie van alle mogelijkheden is dan ook maar mogelijk voor $N \leq 30^3$ [22, 5, 24]. Een polynomiaal algoritme dat een binair perceptron met maximale stabiliteit tot resultaat heeft is tot vandaag de dag onbekend. Het is namelijk zo dat het vinden van een binair perceptron dat alle patronen met een maximale stabiliteit opslaat een NP-compleet probleem is voor alle waarden van $\alpha > 0$ [28, 3]. De NP-compleetheid van het probleem sluit echter niet uit dat er een polynomiaal algoritme bestaat dat een bijna optimale oplossing kan vinden. Er zijn dan ook vele pogingen ondernomen. Ze kunnen allen in twee grote groepen onderverdeeld worden. De eerste soort van strategie is direct werkzaam in de discrete ruimte van de 2^N binaire vectoren. Deze vectoren vormen de hoekpunten van een N -dimensionale hyperkubus. Door gebruik te maken van een kostfunctie die kleine of negatieve waarden van de stabiliteiten benadeeld tracht men een van de hoekpunten van de hyperkubus te vinden met de laagste kost. Het minimaliseren van de kostfunctie gebeurt met technieken zoals "simulated annealing" [13, 18] of genetische algoritmen [18, 7]. De minimalisatie is enorm moeilijk. De kostfunctie heeft een groot aantal lokale minima waarin de minimalisatie kan vast geraken. Desalniettemin heeft de methode in zijn meest gesofistikeerde vorm [18] toch vruchten afgeworpen voor netwerken tot $N = 100$. De resultaten worden echter vlug slechter voor groter wordende waarden van N .

De tweede meer aantrekkelijke aanpak tracht voordeel te halen uit het feit dat er efficiënte algoritmen bestaan voor het probleem met continue koppelingen [26, 31, 27]. Uit deze koppelingen kan men dan trachten de gezochte binaire koppelingen te bepalen. De onderliggende motivering is gebaseerd op de aanname dat het teken van de synapsen van de MSN en de MSB sterk gecorreleerd zijn omdat ze beide dezelfde stabiliteit willen maximaliseren. Daarom moet het

²MSB staat voor Maximaal Stabiel Binair.

³Deze beperking tot $N = 30$ is enkel waar als men een Gray-code algoritme gebruikt als enumeratietechniek. Bij gebruik van een branch en bound techniek is het mogelijk om resultaten te bekomen die equivalent zijn met een enumeratie van $N = 40$ [23]. Ik kom hier later nog op terug.

1. Probleemstelling

mogelijk zijn van nuttige informatie over de componenten van de MSB te halen uit de componenten van de MSN. Het onderwerp van deze thesis handelt over de studie van dergelijke algoritmen en we komen hierop dan ook nog uitgebreid terug in de volgende hoofdstukken.

Een motivering voor binaire koppelingen

Waarom trachten we het associatief geheugenprobleem op te lossen voor binaire koppelingen als er al zeer efficiënte algoritmen bestaan voor continue koppelingen?

Allereerst is er natuurlijk de zuiver wetenschappelijke interesse, vooral met in het achterhoofd het feit dat het een NP-compleet probleem is.

Anderzijds, voor praktische implementaties van dergelijke netwerken, zal men aan discrete koppelingen de voorkeur geven omdat deze een veel eenvoudiger hard-ware realisatie hebben. In het geval de koppelingen enkel de waarden 1 en 0 kunnen aannemen komt dit bij de constructie van het netwerk respectievelijk overeen met het maken van een connectie of niet. Continue koppelingen zijn vanuit praktisch oogpunt minder aantrekkelijk gezien de moeilijkere hard-ware implementatie. Diametraal hierop is de keuze die men zou maken naargelang de beschikbaarheid van een algoritme voor het vinden van de gewichten. Hier zal men dan opteren voor continue koppelingen. In het geval van discrete synapsen is het enige algoritme immers de volledige exploratie van alle mogelijkheden.

Het is dus wenselijk van een strategie te bepalen voor het trainen van netwerken met binaire koppelingen.

1.3 Overzicht

In de volgende hoofdstukken heb ik getracht van de tekst zo weinig mogelijk te overladen met formules. Dit heeft echter wel tot gevolg dat de formules die wel vermeld worden min of meer uit de lucht komen vallen. De afleiding die aan het resultaat vooraf gaat is echter steeds opgenomen in een bijlage achteraan in de thesis. In contrast met het voorgaande heb ik echter wel geopteerd van de zadelpuntvergelijkingen voor de verschillende beschouwde systemen op te nemen in de tekst. Dit zijn natuurlijk geen belangrijke formules, maar het ontbreken ervan maakte het soms moeilijker om te spreken over bepaalde concepten, zoals ordeparameters. Vermits ik een tekst wou die volledig los

1.3. Overzicht

van de bijlagen zou kunnen gelezen worden heb ik de zadelpuntvergelijkingen opgenomen in de tekst.

In **hoofdstuk 2** zal ik het idee van het gebruik van een continu perceptron om componenten van de MSB te voorspellen verder ontwikkelen. Dit hoofdstuk beschrijft de algemene theorie zonder specifiek een keuze te maken van de continue leerregel. **Hoofdstuk 3 en 4** behandelen dan verschillende continue leerregels die gebruikt worden om componenten van de MSB te voorspellen. Hoofdstuk 3 beschouwd leerregels op de hypersfeer en hoofdstuk 4 in de hyperkubus. **Hoofdstuk 5** behandelt een strategie gebaseerd op de voorgaande resultaten om een exacte oplossing te bekomen voor de MSB met systemen tot $N = 50$.

1. Probleemstelling

Hoofdstuk 2

Het precursor idee

2.1 Inleiding

In dit en volgende hoofdstukken gaan we dieper in op het idee om een continue perceptron te gebruiken als voorbode, of precursor, op het vinden van een binair perceptron met maximale stabiliteit. Bedoeling is dan te trachten van de synapsen van de MSB te bepalen uit deze van de continue precursor. Dit idee was reeds vermeld in vorig hoofdstuk waar aangestipt was dat er een correlatie verwacht wordt tussen het teken van de componenten van de MSN en de MSB. We zullen onderzoeken hoeveel informatie over de MSB kan bekomen worden uit een perceptron met continue koppelingen waarin dezelfde set van patronen η^μ opgeslagen is als dewelke men in het binair perceptron wenst op te slaan.

De meest voor de hand liggende methode om via een continue synaps J_i een binaire synaps B_i te bekomen is het zogenaamde "knippen"¹: $B_i = \text{sign}(J_i)$.

Penney en Sherrington [26] hebben voor $N \rightarrow \infty$ de fractie van componenten berekend, van de volledig geknipte MSN, die correct de overeenkomstige component van de MSB voorspellen. Deze fractie is groot. Zij vonden dat in het slechtste geval, bij $\alpha = 0.83$, 80% van de componenten correct voorspeld zijn. Dit percentage neemt toe tot ongeveer 90% bij $\alpha = 0.1$. Desondanks dit hoge percentage van correct geknipte componenten is een volledig knippen van de MSN geen goede benadering voor de MSB. Dit blijkt uit de distributie van de stabiliteiten van de geknipte MSN-vektor. Penney en Sherrington vonden

¹In de Engelse literatuur wordt dit "clipping" genoemd. Ik heb hier voor het woord "knippen" gekozen in de zin dat het teken van de continue component "losgeknipt" wordt van zijn absolute getalwaarde en toegekend wordt aan de overeenkomstige binaire component.

2. Het precursor idee

dat bij $\alpha = 0.8$ ongeveer 17% van de stabiliteiten negatief zijn. Dit aantal vermindert wel tot 5% bij $\alpha = 0.1$, maar is toch een indicatie dat volledig knippen niet voldoende is. Indien men echter de componenten die correct de overeenkomstige component van de MSB voorspellen kan identificeren, zou men deze kunnen knippen en de andere onderwerpen aan een tweede leerproces om zo een betere benadering te bekomen van de MSB. Op basis van numerieke simulaties voor kleine systemen $N \leq 25$, hebben Penny en Sherrington voorgesteld dat de componenten van de MSN met de grootste getalwaarde zeer waarschijnlijk de overeenkomstige component van de MSB voorspellen. Dit betekent dat de 20% foute tekens in de geknipte MSN voornamelijk moeten gezocht worden onder de 20 tot 40% kleinste² synapsen van de MSN. Als deze suggestie correct is voor alle waarden van N reduceert de effectieve dimensie van het probleem vermits 60% van de synapsen onmiddellijk kunnen bekomen worden uit de MSN. De andere componenten moeten dan met een aanvullend algoritme bepaald worden.

In deze thesis zal ik naast de MSN ook andere continue leerregels beschouwen en onderzoeken hoe goed ze bij knippen van hun componenten de taak van precursor voor de MSB vervullen. In het bijzonder wensen we te weten hoeveel componenten bij knippen correct de overeenkomstige component van de MSB voorspellen. Tevens wensen we deze componenten te kunnen identificeren. Hiertoe zal ik de kwalitatieve suggestie van Penney en Sherrington kwantitatief maken voor de MSN en andere continue leerregels. Dit zal ik doen door te berekenen wat de fractie is van correct geknipte componenten als we enkel een zeker aantal van de grootste componenten van de precursor knippen.

2.2 Kwaliteitsmaat voor een continue precursor

Om verschillende precursors kwalitatief met mekaar te kunnen vergelijken hebben we een kwaliteitsmaat nodig. Er zijn verschillende grootheden die hiervoor kunnen gebruikt worden, zoals de overlap tussen de continue precursor koppelingen $\mathbf{J}$ en de MSB-vektor $\mathbf{B}$. Des te groter deze overlap des te beter de precursor zal zijn. De keuze die Penney en Sherrington genomen hebben is de

²Wanneer in deze context gesproken wordt over de grootste en kleinste componenten wordt dit steeds bedoeld t. o. v. de getalwaarde van de synapsen. Dus de componenten groter dan $J_o > 0$ voldoen aan $|J| > J_o$.

2.3. De gemeenschappelijke distributie $P(J, B)$

overlap tussen de geknipte precursor $\text{sign}(\mathbf{J})$ en³ $\mathbf{B}$, onder de volgende vorm:

$$f_1 \equiv \frac{1}{2} \left(1 + \frac{1}{N} \text{sign}(\mathbf{J}) \cdot \mathbf{B} \right). \quad (2.1)$$

De grootheid f_1 is dus de fractie van binaire koppelingen die correct kunnen bekomen worden door het volledig knippen van het precursor perceptron. In het licht van de in de laatste paragraaf van vorige sectie 2.1 vermelde doelstellingen is het echter gewenst een veralgemening in te voeren van f_1 . Immers, f_1 laat niet toe te berekenen wat de fractie van correct geknipte componenten is als enkel een aantal van de grootste componenten geknipt worden. We definiëren dan ook volgende kwaliteitsmaat:

Definitie *De kwaliteitsmaat voor een continue precursor voor de MSB noteren we als f_x met $x \in [0, 1]$. Het geeft de fractie van correct geknipte componenten als enkel de xN grootste componenten (in absolute waarde) van de precursor geknipt worden.*

Het is onmiddellijk duidelijk dat voor $x = 1$ de kwaliteitsmaat (2.1) bekomen wordt. Deze grootheid f_x kan eenvoudig berekend worden als we de gemeenschappelijke waarschijnlijkheidsdistributie $P(J, B)$ kennen van twee overeenkomstige componenten in de synaps vektoren $\mathbf{J}$ en $\mathbf{B}$. Om wille van symmetrieredenen is $P(J, B) = P(-J, -B)$ en dus geldt er voor f_x :

$$f_x = \frac{2}{x} \int_{J_o(x)}^{+\infty} dJ P(J, B = 1), \quad (2.2)$$

waarbij $J_o(x)$ bepaald wordt uit $x = 2 \int_{J_o}^{+\infty} dJ P(J)$.

2.3 De gemeenschappelijke distributie $P(J, B)$

Om de grootheid $P(J, B)$ uit (2.2) te berekenen zal ik de werkwijze van Wong, Rau en Sherrington volgen [34]. Deze bestaat erin het gecombineerde systeem van de continue precursor en het MSB perceptron te beschouwen voor dezelfde set van patronen $\boldsymbol{\eta}^\mu$. Deze berekening wordt uitgevoerd met behulp van de replica methode in de thermodynamische limiet $N \rightarrow \infty$ en $\alpha = p/N$ eindig. Ik zal voorlopig in deze berekening de leerregel voor de precursor zo weinig mogelijk specificeren. Achteraf kan dan een specifieke leerregel ingevuld worden.

³Met $\text{sign}(\mathbf{J})$ wordt bedoeld $(\text{sign}(J_1), \text{sign}(J_2), \dots, \text{sign}(J_N))$

2. Het precursor idee

De mogelijke precursor leerregels die ik zal beschouwen behoren tot een klasse waarvan de kostfuncties allemaal zijn van de vorm $E(\mathbf{J}) = \sum_{\mu} V(\lambda_{\mu})$ met $\lambda_{\mu} = \frac{1}{\sqrt{N}} \mathbf{J} \cdot \boldsymbol{\eta}^{\mu}$ [4]. De precursor $\mathbf{J}^*$ wordt dan bepaald door:

$$\mathbf{J}^* = \text{Arg Min}_{\mathbf{J}} \left[\sum_{\mu} V(\lambda_{\mu}) \right] \quad \text{met } \lambda_{\mu} = \frac{1}{\sqrt{N}} \mathbf{J} \cdot \boldsymbol{\eta}^{\mu}. \quad (2.3)$$

Hier zijn de p patronen $\boldsymbol{\eta}^{\mu}$ willekeurig gekozen uit een distributie die voldoet aan (1.8). De leerregel is pas volledig gespecificeerd als naast de potentiaal V ook het gebied waarop de minimalisatie moet uitgevoerd worden gegeven is. Mogelijke minimalisatiegebieden zullen mathematisch ingevoerd worden door een waarschijnlijkheidsdistributie $p(\mathbf{J})$ die binnen het minimalisatiegebied een constante waarde aanneemt en nul is daarbuiten.

Veel gebruikte leerregels zijn vervat in deze klasse van algoritmen door een geschikte keuze te maken van de potentiaal $V(\lambda)$ en $p(\mathbf{J})$. Enkele voorbeelden zijn:

$V(\lambda)$	$p(\mathbf{J})$	Leerregel	Referentie
$-\lambda$	$p(\mathbf{J}) = \delta(\mathbf{J}^2 - N)$	Hebb	[11]
$(\kappa - \lambda)^2 \Theta(\kappa - \lambda)$	$p(\mathbf{J}) = \delta(\mathbf{J}^2 - N)$	AdaTron	[11]
$\frac{1}{\lambda - \kappa}$	$p(\mathbf{J}) = \delta(\mathbf{J}^2 - N)$	hypersfeer	zie Hfst. 3
$\frac{1}{\lambda - \kappa}$	$\prod_i \frac{1}{2} \Theta(1 - J_i^2)$	hyperkubus	zie Hfst. 4

In wat volgt zal ik aannemen dat de distributie $p(\mathbf{J})$ geschreven kan worden onder de vorm $p(\mathbf{J}) = \prod_i p(J_i)$, zoals voor de laatste leerregel in de tabel het geval is. Voor de eerste drie leerregels in de tabel is de distributie die het minimalisatiedomein definieert een *globale* constraint, $\mathbf{J}^2 = N$. Ze verschilt van de laatste doordat de normeringsvoorwaarde een verband tussen alle componenten van $\mathbf{J}$ is en niet een aparte voorwaarde voor elke component. Ze kunnen dus niet geschreven worden onder de vorm die we wensen voorop te stellen: $p(\mathbf{J}) = \prod_i p(J_i)$. Ik zal later echter laten zien hoe de resultaten van de berekening eenvoudig aangepast kunnen worden tot dergelijke globale constraints.

2.3. De gemeenschappelijke distributie $P(J, B)$

De MSB leerregel zullen we ook uitdrukken m. b. v. een kostfunctie binnen dezelfde klasse als de precursor leerregel:

$$B^* = \text{Arg Min}_B \left[\sum_{\mu} \Phi(\Lambda_{\mu}) \right] \quad \text{met } \Lambda_{\mu} = \frac{1}{\sqrt{N}} B \cdot \eta^{\mu}, \quad (2.4)$$

hierbij is [20]:

$$\begin{aligned} \Phi(\Lambda) &= \Theta(\kappa_b - \Lambda) \\ p(B_i) &= \frac{1}{2} [\delta(B_i + 1) + \delta(B_i - 1)], \end{aligned} \quad (2.5)$$

waarbij de laatste uitdrukking het minimalisatiedomein definieert als de hoekpunten van de N -dimensionale hyperkubus. De stabiliteit κ_b heeft een sub-index b gekregen om te benadrukken dat het de stabiliteit is van het binair perceptron. Verder zijn we niet zomaar tevreden met gelijk welk binair perceptron dat de kostfunctie minimaliseert. We willen de MSB als oplossing, die de patronen stockeert met maximale stabiliteit. Om deze oplossing te bekomen zullen we dus de leerregel moeten beschouwen bij verzadiging van het perceptron, dus bij $\alpha = \alpha_c(\kappa_b)$. Ik zal verder in dit hoofdstuk bespreken hoe deze voorwaarde mathematisch kan opgelegd worden.

Met bovenstaande notatie schrijven we nu

$$\begin{aligned} P(J, B) &= f(J) g(B) \\ &= \prod_i p(J_i) \frac{e^{-\beta \sum_{\mu} V(\Lambda_{\mu})}}{Z_J(\beta)} \prod_i p(B_i) \frac{e^{-\beta \sum_{\mu} \Phi(\Lambda_{\mu})}}{Z_B(\beta)}, \end{aligned} \quad (2.6)$$

voor de vektoren J en B . De waarschijnlijkheidsdistributie is dus gegeven door een produkt van twee Gibbs distributies met temperatuur $T = 1/\beta$ en is gedefinieerd op het gekozen minimalisatiedomein bepaalt door de distributies $p(J_i)$ en $p(B_i)$ voor de twee leerregels afzonderlijk. Als we de limiet nemen voor $\beta \rightarrow \infty$ zullen enkel de perceptrons die de kostfuncties V en Φ optimaliseren in beschouwing worden genomen. Resultaten voor $\beta \rightarrow \infty$ hebben dan ook alleen maar betrekking op de optimale perceptrons J^* en B^* , wat gewenst is.

Wat we nu willen berekenen is $P(J_i, B_i)$ voor een willekeurige component i en voor een specifiek set van patronen. Deze distributie wordt aangenomen van "self-averaging" te zijn in de limiet voor $N \rightarrow \infty$. Dit betekent dat de

2. Het precursor idee

distributie voor $N \rightarrow \infty$ steeds dezelfde is wat ook de keuze van de patronen is. Men kan dus $P(J_i, B_i)$ middelen over de patronen zonder het resultaat bij $N \rightarrow \infty$ te veranderen. Deze middeling maakt het mogelijk van $P(J_i, B_i)$ te berekenen. We kunnen $P(J_i, B_i)$ dan schrijven als:

$$P(j, b) = \langle \int d\mathbf{J} d\mathbf{B} \delta(J_1 - j) \delta(B_1 - b) P(\mathbf{J}, \mathbf{B}) \rangle. \quad (2.7)$$

Hier geeft $\langle \rangle$ de middeling weer over de patronen $\boldsymbol{\eta}^\mu$. Vermits deze patronen willekeurig gekozen zijn is $P(J_i, B_i)$ ook onafhankelijk van de gekozen component i . In (2.7) is dan ook arbitrair $i = 1$ gekozen.

We zullen deze uitdrukking voor $P(j, b)$ interpreteren als de kans dat een willekeurige component J_i en een willekeurige component B_i respectievelijk de waarden j en b aannemen bij een specifieke set van patronen.

2.3.1 De keuze van de leerregels

Zonder hier reeds expliciete keuzes te maken voor $V(\lambda)$ en $p(J_i)$ zal ik hier echter wel de mogelijkheden beperken. Ik zal eisen dat de potentiaal $V(\lambda)$ en $p(J_i)$ zodanig zijn dat de kostfunctie (2.3) één uniek minimum heeft. De reden hiervoor is tweërlei. Ten eerste is er een mathematische reden: ik zal $P(j, b)$ in de volgende sectie berekenen met de replica symmetrie-ansatz (RS-ansatz). Aanname van het unieke minimum van de kostfunctie verzekert een correcte oplossing onder deze ansatz⁴. Ten tweede is er een meer praktische reden. Indien de kostfunctie geen uniek minimum zou hebben is het numeriek bepalen van het absolute minimum en dus het vinden van de precursor onbegonnen werk. De numerieke minimalisatie zou steeds vast geraken in lokale minima, zodat tevens de kwaliteit van de precursor zou afhangen van de dynamika van het minimalisatieproces. Vermits het vinden van de MSB al een ontzagwekkende taak is zou het onverstandig zijn de precursor strategieodeloos te verzwaren.

2.4 Resultaten voor $P(j, b)$

De grootheid $P(j, b)$ is berekend voor $N \rightarrow \infty$ in Bijlage A op bladzijde 77 e. v. onder de aanname van replica symmetrie. Hierbij is tevens de limiet voor

⁴Het blijkt dat de binaire leerregel (2.5) bij verzadiging van het netwerk, nog steeds geen uniek minimum heeft. Krauth en Mézard hebben echter getoond [20] dat de RS-vergelijkingen nog steeds geldig zijn. Zie verder in de tekst.

2.4. Resultaten voor $P(j, b)$

$\beta \rightarrow \infty$ genomen om resultaten te bekomen die enkel betrekking hebben op de perceptrons die de kostfunctie minimaliseren. Ik zal hier enkel de resultaten vermelden. De volgende ordeparameters komen naar voren in de berekening:

$$\begin{aligned} f_a &= \frac{1}{N} \sum_j (J_j^a)^2; & r_{ab} &= \frac{1}{N} \sum_j J_j^a B_j^b \\ q_{ab} &= \frac{1}{N} \sum_j J_j^a J_j^b; & Q_{ab} &= \frac{1}{N} \sum_j B_j^a B_j^b, \end{aligned} \quad (2.8)$$

en dit natuurlijk samen met hun toegevoegden $\hat{f}, \hat{q}, \hat{Q}$ en $\hat{r}$. De parameters $f, \hat{f}$ en $q, \hat{q}$ (of Q en $\hat{Q}$ voor het MSB systeem) zijn de gebruikelijke ordeparameters van elk systeem afzonderlijk⁵. Hier is f de zelf overlap en q de overlap tussen twee perceptrons uit verschillende replicas. Verder zijn er twee nieuwe ordeparameters, namelijk de overlap r tussen de perceptrons van de twee afzonderlijke systemen en zijn toegevoegde $\hat{r}$. De waarde van de ordeparameters wordt bepaald door hun zadelpuntvergelijkingen. In de volgende drie secties zal ik deze vergelijkingen specificeren. Voor details betreffende de berekeningen verwijs ik naar bijlage A.

2.4.1 De zadelpuntvergelijkingen voor het precursor systeem

Omdat de potentiaal $V(\lambda)$ een uniek minimum heeft volgt er voor $\beta \rightarrow \infty$ dat er maar één perceptron overblijft. Dit impliceert dat voor $\beta \rightarrow \infty$, $q \rightarrow f$ [10]. Met de invoering van $x = \beta(f - q)$ zijn de zadelpuntvergelijkingen gegeven door

$$(2\hat{f} + \hat{q}) + \frac{\alpha}{(f - q)\sqrt{f}} \int_{-\infty}^{+\infty} D\sigma \sigma \left[\lambda_o(\sqrt{f}\sigma, x) - \sqrt{f}\sigma \right] = 0, \quad (2.9)$$

$$\hat{f} - \frac{\alpha}{2(f - q)} + \frac{\alpha}{2(f - q)^2} \int_{-\infty}^{+\infty} D\sigma \left[\lambda_o(\sqrt{f}\sigma, x) - \sqrt{f}\sigma \right]^2 = 0, \quad (2.10)$$

$$-\frac{1}{2}(f - q) + \int_{-\infty}^{+\infty} D\varphi \frac{\int_{-\infty}^{+\infty} d\mu(J) e^{-\frac{1}{2}(2\hat{f} + \hat{q})J^2 + \sqrt{\hat{q}}J\varphi} \frac{J\varphi}{2\sqrt{\hat{q}}}}{\int_{-\infty}^{+\infty} d\mu(J) e^{-\frac{1}{2}(2\hat{f} + \hat{q})J^2 + \sqrt{\hat{q}}J\varphi}} = 0, \quad (2.11)$$

⁵Bemerk dat ik de ordeparameters voor het MSB systeem in hoofdletters noteer en die van het precursor systeem in kleine letters.

2. Het precursor idee

$$\frac{1}{2}f + \int_{-\infty}^{+\infty} D\varphi \frac{\int_{-\infty}^{+\infty} d\mu(J) e^{-\frac{1}{2}(2\hat{f}+\hat{q})J^2 + \sqrt{\hat{q}}J\varphi} \cdot \left(-\frac{1}{2}J^2\right)}{\int_{-\infty}^{+\infty} d\mu(J) e^{-\frac{1}{2}(2\hat{f}+\hat{q})J^2 + \sqrt{\hat{q}}J\varphi}} = 0. \quad (2.12)$$

Hierin is λ_o gegeven door:

$$\lambda_o(\sqrt{f}\sigma, x) = \text{Arg Min}_{\lambda} \left[V(\lambda) + \frac{(\lambda - \sqrt{f}\sigma)^2}{2x} \right]. \quad (2.13)$$

Merk op dat men nog wel beperkt is in het uitrekenen van de vergelijkingen omdat de distributie $p(J)$ en de potentiaal $V(\lambda)$ niet expliciet gekend zijn. Een verdere studie van de vergelijkingen is dan ook geval per geval nodig. In de volgende hoofdstukken zal ik specifieke keuzes voor V en $p(J_i)$ invullen.

2.4.2 De zadelpuntvergelijkingen voor het MSB systeem

Er is vroeger reeds opgemerkt dat de leerregel (2.5) moet toegepast worden bij verzadiging van het netwerk om de MSB als oplossing te bekomen. We willen dus een oplossing van de zadelpuntvergelijkingen bij verzadiging van het netwerk omdat deze dan de ordeparameters van de MSB als oplossing hebben. Bij een bepaalde waarde van κ_b zijn er voor $\alpha < \alpha_c(\kappa_b)$ over het algemeen vele binaire perceptrons die de kostfunctie bepaalt door $\Phi(\Lambda)$ minimaliseren. Verwacht wordt dat dit aantal oplossingen kleiner en kleiner wordt totdat, bij $\alpha = \alpha_c(\kappa_b)$, er nog 1 oplossing overblijft. In deze limiet van verzadiging van het netwerk verwacht men dan ook dat $Q \rightarrow 1$. Dit is nochtans fout. Gardner en Derrida bekomen met de aanname dat voor $\beta \rightarrow \infty$, $Q \rightarrow 1$ een geheugencapaciteit $\alpha_c = \frac{4}{\pi}$ bij $\kappa_b = 0$ [10]. De replica symmetrie entropie S_{RS} is echter $-\infty$ bij deze oplossing. Dit is zeker fout voor een binair systeem vermits het een discreet aantal toestanden heeft. Krauth en Mézard hebben echter aangetoond[20] dat de replica symmetrie berekening toch juist⁶ is zolang de replica symmetrie entropie $S_{RS} \geq 0$. Er geldt nu dat, bij κ_b vast, S_{RS} een strikt monotoon dalende functie is van α . Zolang de entropie $S_{RS} > 0$ zijn er exponentieel veel binaire perceptrons die de kostfunctie minimaliseren maar zodra $S_{RS} = 0$ zijn er dat plots maar polynomiaal veel. De waarde van α waarop

⁶Meer exact: de resultaten van de RS berekening zijn dezelfde als de resultaten bij een 1stap-RSB berekening.

2.4. Resultaten voor $P(j, b)$

dit gebeurt blijkt de juiste geheugencapaciteit te zijn van het perceptron [20]. (zie figuur 2.2) De zadelpuntvergelijkingen oplossen met de extra voorwaarde $S_{RS} = 0$ levert dan ook de ordeparameters van de MSB. Bij $\kappa_b = 0$ blijkt in dit geval dat $\alpha_c = 0.83$ en dat $Q = 0.56$ en dus niet gelijk is aan 1 wat duidelijk suggereert dat de MSB geen unieke vektor is. Vermits het onmogelijk is onderscheid te maken tussen al deze vectoren hebben de theoretische resultaten betrekking op gemiddelden over dit ensemble van vectoren. Dit geldt dan ook voor $P(j, b)$. Dus alle algoritmen die we construeren op basis van de grootheid $P(j, b)$ om de MSB te bepalen zullen op zijn best een binaire vektor opleveren die gericht is naar het gemiddelde $\langle \mathbf{B} \rangle$ over dit ensemble van vectoren en niet naar een specieke vektor $\mathbf{B}$.

De expliciete vorm van de zadelpuntvergelijkingen wordt gegeven door:

$$1 - Q = \int_{-\infty}^{+\infty} D\varphi \frac{1}{\cosh^2(\sqrt{\hat{Q}}\varphi)}, \quad (2.14)$$

$$\hat{Q} = \frac{\alpha}{2\pi(1-Q)} \int_{-\infty}^{+\infty} D\sigma \frac{e^{-\frac{(\kappa_b - \sqrt{Q}\sigma)^2}{1-Q}}}{H\left(\frac{\kappa_b - \sqrt{Q}\sigma}{\sqrt{1-Q}}\right)}. \quad (2.15)$$

En de "zero-entropie" voorwaarde wordt gegeven door:

$$S_{RS} = \frac{Q\hat{Q}}{2} + \alpha \int_{-\infty}^{+\infty} Dz \ln H\left(\frac{\kappa_b + \sqrt{Q}z}{\sqrt{1-Q}}\right) + \int_{-\infty}^{+\infty} Dz \ln \left[e^{-\frac{\hat{Q}}{2}} 2 \cosh\left(\sqrt{\hat{Q}}z\right) \right] = 0. \quad (2.16)$$

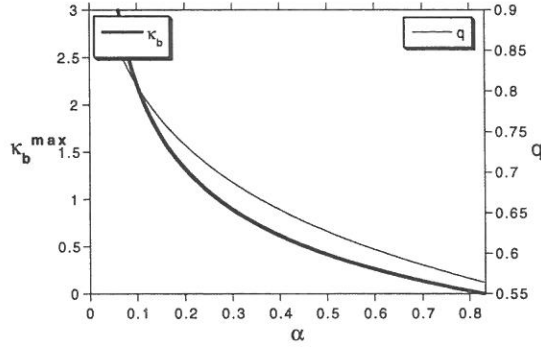
De numerieke oplossing van het bovenstaande stelsel vergelijkingen is gegeven in figuur (2.1). Figuur 2.2 geeft de entropie $S_{RS}(\alpha)$ bij een aantal waarden van κ_b . De MSB oplossing zijn aangeduid met pijltjes.

2.4.3 De zadelpuntvergelijkingen voor r en $\hat{r}$

Tenslotte zijn de zadelpuntvergelijkingen voor r en $\hat{r}$ gegeven door:

$$r = \int_{-\infty}^{+\infty} D\hat{\gamma}(\varphi, \psi) \cdot \tanh(\sqrt{\hat{Q}}\psi) \cdot \frac{\int_{-\infty}^{+\infty} d\mu(J) J \exp\left[-\frac{1}{2}(2\hat{f} + \hat{q})J^2 + \sqrt{\hat{q}}J\varphi\right]}{\int_{-\infty}^{+\infty} d\mu(J) \exp\left[-\frac{1}{2}(2\hat{f} + \hat{q})J^2 + \sqrt{\hat{q}}J\varphi\right]}, \quad (2.17)$$

2. Het precursor idee



Figuur 2.1: Numerieke oplossing van de MSB vergelijkingen. De dikke lijn geeft $\kappa_b^{max}(\alpha)$ en de dunne lijn geeft $q(\alpha)$, welke geplot is op de rechter y-as. Bemerk dat bij $\alpha = 0.83$, $q \approx 0.56$.

$$\hat{r} = \alpha \int_{-\infty}^{+\infty} D_\gamma(\tau, \sigma) \cdot \frac{\lambda_o - \sqrt{q}\sigma}{f - q} \cdot \frac{1}{\sqrt{2\pi(1-Q)}} \cdot e^{-\frac{(\kappa_b - \sqrt{Q}\tau)^2}{2(1-Q)}} \cdot \frac{1}{H\left(\frac{\kappa_b - \sqrt{Q}\tau}{\sqrt{1-Q}}\right)} d\tau. \quad (2.18)$$

Hierbij is λ_o gegeven door (2.13). Verder is $D_\gamma(\tau, \sigma)$ een verkorte notatie voor de Gaussische integratiemaat met correlatie γ (E.3). De parameters γ en $\hat{\gamma}$ staan symbool voor:

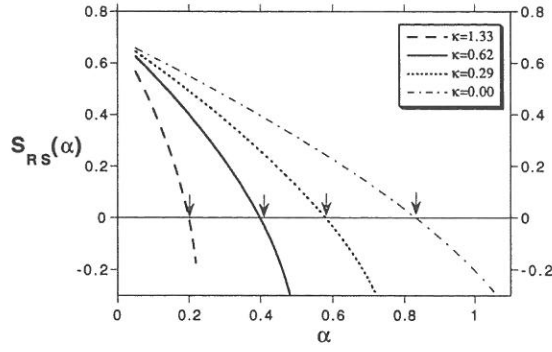
$$\gamma = \frac{r}{\sqrt{qQ}}, \quad \hat{\gamma} = \frac{\hat{r}}{\sqrt{\hat{q}\hat{Q}}}, \quad (2.19)$$

en zijn dus equivalent met r en $\hat{r}$. In deze vergelijkingen dienen tevens de juiste keuze van potentiaal en distributie p ingevuld te worden.

2.4.4 De uitdrukking voor $P(j, b)$

De set zadelpuntvergelijkingen voor elk systeem afzonderlijk zijn niet gekoppeld met elkaar alsook niet met de set vergelijkingen voor r en $\hat{r}$. Elke set kan dus afzonderlijk opgelost worden. Vermits de vergelijkingen voor r en $\hat{r}$ echter de ordeparameters van de afzonderlijke systemen bevatten moeten eerst die vergelijkingen opgelost worden waarna de twee zadelpuntvergelijkingen voor r

2.4. Resultaten voor $P(j, b)$



Figuur 2.2: De replica symmetrie entropie $S_{RS}(\alpha)$ bij een aantal waarden van κ_b . De oplossing bij verzadiging, de MSB, is telkens aangeduid met een pijltje. De entropie heeft daar de waarde nul.

en $\hat{r}$ kunnen worden opgelost voor specifieke waarden van de andere ordeparameters.

Zodra de ordeparameters voor het gehele systeem bekend zijn kan $P(j, b)$ berekend worden als volgt:

$$P(j, b) = \int_{-\infty}^{+\infty} D_{\hat{\gamma}}(\varphi, \psi) \mathcal{J}(\varphi) \mathcal{B}(\psi), \quad (2.20)$$

hierin wordt $\mathcal{J}$ gegeven door:

$$\mathcal{J}(\varphi) = \frac{p_J(j) e^{-\frac{1}{2}(2\hat{f} + \hat{q})(j - \frac{\sqrt{\hat{q}}}{2\hat{f} + \hat{q}}\varphi)^2}}{\int_{-\infty}^{+\infty} d\mu(J) \cdot e^{-\frac{1}{2}(2\hat{f} + \hat{q})(J - \frac{\sqrt{\hat{q}}}{2\hat{f} + \hat{q}}\varphi)^2}}, \quad (2.21)$$

2. Het precursor idee

en $\mathcal{B}$ door:

$$\mathcal{B}(\psi) = \frac{p_B(b) e^{-\frac{\hat{Q}}{2}(b - \frac{\psi}{\sqrt{\hat{Q}}})^2}}{\int_{-\infty}^{+\infty} d\mu(B) \cdot e^{-\frac{\hat{Q}}{2}(B - \frac{\psi}{\sqrt{\hat{Q}}})^2}}. \quad (2.22)$$

Hierbij wordt met $p_J(j)$ bedoeld de distributie die het minimalisatiedomein voor het $\mathbf{J}$ -systeem bepaalt, geëvalueerd in j . Analoog voor $p_B(b)$.

Wat onder globale constraints?

De zadelpuntvergelijkingen en de uitdrukking voor $P(j, b)$ kunnen zeer eenvoudig aangepast worden aan globale constraints. Veronderstel dat het $\mathbf{J}$ systeem een minimalisatiedomein heeft bepaalt door $P(\mathbf{J}) = \delta(\mathbf{J}^2 - N)$ dan volstaat het uitvoeren van de volgende transformatie om de vergelijkingen aan te passen:

$$\begin{aligned} f &= 1 \\ \int d\mu(J) &= \int dJ, \end{aligned} \quad (2.23)$$

waarbij tevens de extrematie naar f natuurlijk moet weggelaten worden. Dit is getoond in Bijlage A op bladzijde 91.

Een woordje uitleg over γ en $\hat{\gamma}$

De parameters $\gamma = r/\sqrt{q\hat{Q}}$ en $\hat{\gamma} = \hat{r}/\sqrt{\hat{q}\hat{Q}}$ hebben een voordeel over r en $\hat{r}$. De ordeparameter r is de overlap tussen het precursor perceptron $\mathbf{J}$ en $\langle \mathbf{B} \rangle$, het zwaartepunt van het ensemble van MSB vectoren.

$$r = \frac{1}{N} \mathbf{J} \cdot \langle \mathbf{B} \rangle = \frac{1}{N} \sqrt{\mathbf{J}^2} \sqrt{\langle \mathbf{B} \rangle^2} \cos \phi, \quad (2.24)$$

hierbij is ϕ de hoek tussen de vectoren $\mathbf{J}$ en $\langle \mathbf{B} \rangle$ en is $\langle \mathbf{B} \rangle$ gegeven door:

$$\langle \mathbf{B} \rangle = \left(\sum_{\mathbf{B}} \mathbf{B} \right) / \left(\sum_{\mathbf{B}} 1 \right), \quad (2.25)$$

waarbij de som loopt over het ensemble van MSB vectoren. Als we $\mathcal{N} = \sum_{\mathbf{B}} 1$ noteren dan kunnen we schrijven voor $\langle \mathbf{B} \rangle^2$:

2.4. Resultaten voor $P(j, b)$

$$\langle \mathbf{B} \rangle^2 = N^{-2} \left(\sum_{\mathbf{B}} \mathbf{B} \right) \cdot \left(\sum_{\mathbf{B}'} \mathbf{B}' \right) = NQ. \quad (2.26)$$

De laatste gelijkheid volgt uit het feit dat twee MSB vektoren uit het ensemble overlap Q hebben. (Zie (2.8) voor de betekenis van de ordeparameter Q .) Uit (2.8) volgt ook dat:

$$\mathbf{J}^2 = Nf, \quad (2.27)$$

zodat men dan (2.24) kan herschrijven tot:

$$r = \frac{1}{N} \sqrt{fN} \cdot \sqrt{NQ} \cdot \cos \phi = \sqrt{fQ} \cos \phi. \quad (2.28)$$

Hieruit volgt dan dat de parameter γ eigenlijk niets anders is dan de cosinus van de hoek tussen $\mathbf{J}$ en $\langle \mathbf{B} \rangle$. De factor $\sqrt{qQ}$ ($q=f$) corrigeert dus voor de lengte van zowel $\mathbf{J}$ als $\langle \mathbf{B} \rangle$.

Terwijl $\hat{r}$ en $\hat{q}$ naar oneindig gaan als $\beta \rightarrow \infty$ blijft $\hat{\gamma}$ een eindige waarde aannemen. Als we $\hat{\gamma}$ in de limiet nemen naar zijn maximale waarde 1 zal $\mathbf{J}$ en $\langle \mathbf{B} \rangle$ dicht bij mekaar komen zodat ook γ naar zijn maximale waarde van 1 zal gaan.

2. Het precursor idee

Hoofdstuk 3

Leren op de hypersfeer

3.1 Inleiding

We zijn nu klaar om de studie van verschillende precursors voor de MSB aan te vatten. De keuze van een precursor bestaat erin de potentiaal $V(\lambda)$ alsook het minimalisatiedomein te specificeren. Zoals reeds vroeger aangestipt is de MSN een evidente keuze als precursor vermits deze dezelfde vergelijkingen oplost als de MSB maar dan in een continue ruimte op de hypersfeer $\mathbf{J}^2 = N$. Het is dan ook vanzelfsprekend het onderzoek te beginnen op de hypersfeer. Men kan zich de vraag stellen of de MSN de meest optimale¹ precursor is. Ik zal in dit hoofdstuk aantonen dat dit niet het geval is. Ik zal afleiden wat het grootste aantal componenten is van de MSB die correct kunnen voorspelt worden met een precursor binnen de klasse van leerregels onder beschouwing. Ik zal tevens een eenvoudige potentiaal voorstellen die een zeer goede benadering oplevert voor de optimale precursor synapsen.

Beschouw een potentiaal $V(\lambda)$ met een uniek minimum op de hypersfeer.

De distributie die het minimalisatiedomein definieert is in dit geval een globale constraint:

$$P(\mathbf{J}) = \delta(\mathbf{J}^2 - N), \quad (3.1)$$

de opmerking van sectie 2.4.4 op bladzijde 24 is dus van toepassing.

¹Met optimaal wordt bedoeld: optimaal in de zin van de gedefinieerde kwaliteitsmaat (2.2).

3. Leren op de hypersfeer

3.2 Het $V(\lambda)$ /MSB systeem

Ik zal hier de zadelpuntvergelijkingen voor het systeem van de MSB en de willekeurige potentiaal op de hypersfeer, alsook de uiteindelijk gewenste kwaliteitsmaat f_x , beschrijven. Voor details betreffende de berekeningen verwijs ik naar Bijlage B.

De zadelpuntvergelijkingen voor de twee systemen afzonderlijk

Vermits de precursor vektor genormeerd is op de hypersfeer $\mathbf{J}^2 = N$, volgt er dat $f = \frac{1}{N}\mathbf{J}^2 = 1$. De zadelpuntvergelijkingen voor het precursor systeem op de hypersfeer reduceren zich dan tot 1 vergelijking voor de parameter $x = \beta(1-q)$:

$$1 = \alpha \int_{-\infty}^{+\infty} D\sigma [\lambda_o(\sigma, x) - \sigma]^2. \quad (3.2)$$

Hierbij is $\lambda_o(\sigma, x)$ weerom bepaald door:

$$\lambda_o(\sigma, x) = \text{Arg Min}_{\lambda} \left[V(\lambda) + \frac{(\lambda - \sigma)^2}{2x} \right]. \quad (3.3)$$

De vergelijkingen voor de MSB zijn reeds gegeven in hoofdstuk 2 :(2.14–2.16).

De zadelpuntvergelijkingen voor r en $\hat{r}$

De zadelpuntvergelijking voor r is gegeven door:

$$r = \alpha \int_{-\infty}^{+\infty} D\sigma [\lambda_o(\sigma, x) - \sigma] \cdot g_r(\sigma), \quad (3.4)$$

met $g_r(\sigma)$ van de volgende vorm:

$$g_r(\sigma) = \sqrt{1-Q} \int_{-\infty}^{+\infty} \frac{d\tau}{\sqrt{2\pi(1-\gamma^2)}} \cdot e^{-\frac{1}{2(1-\gamma^2)}[\tau-\gamma\sigma]^2} \cdot \frac{1}{\sqrt{2\pi}} \frac{e^{-\frac{(\kappa_b - \sqrt{Q}\tau)^2}{2(1-Q)}}}{H\left(\frac{\kappa_b - \sqrt{Q}\tau}{\sqrt{1-Q}}\right)}. \quad (3.5)$$

Bemerk dat de afhankelijkheid van de potentiaal $V(\lambda)$ enkel optreedt in de eerste term $(\lambda_o - \sigma)$ terwijl de tweede factor $g_r(\sigma)$ grotendeels bepaald is door de MSB ordeparameter Q en door κ_b . De r afhankelijkheid is aanwezig via de parameter $\gamma = r/\sqrt{Q}$. Tenslotte wordt $\hat{r}$ bepaald door:

3.2. Het $V(\lambda)$ /MSB systeem

$$r = \frac{\hat{r}}{\sqrt{\hat{q}}}(1 - Q). \quad (3.6)$$

Zoals reeds vermeld in hoofdstuk 2 moeten dus eerst de ordeparameters van beide systemen (precursor en MSB) bepaald worden—als functie van α —uit hun respectievelijke zadelpuntvergelijkingen. Vervolgens dienen r en $\hat{r}$ berekend te worden als functie van α . Zodra de volledige set van ordeparameters gekend is kan men overgaan tot het berekenen van de kwaliteitsmaat f_x m. b. v. de gevonden uitdrukking voor $P(j, b)$.

Uitdrukking voor de kwaliteitsmaat f_x

De distributie $P(j, b)$ bepaald door de algemene uitdrukking (2.20...2.22) wordt nu gegeven door:

$$P(j, b) = \int_{-\infty}^{+\infty} D_{\hat{\gamma}}(\varphi, \psi) \delta(j - \varphi) \frac{1}{2} [\delta(b + 1) + \delta(b - 1)] \cdot \frac{e^{\sqrt{\hat{Q}}b\psi}}{\cosh(\sqrt{\hat{Q}}\psi)}. \quad (3.7)$$

Met deze uitdrukking voor $P(j, b)$ kan men nu de fractie van componenten $|J| > J_o > 0$ berekenen die bij knippen correct de corresponderende component van de MSB opleveren. Vermits $P(j, b) = P(-j, -b)$ moeten we enkel rekening houden met bijvoorbeeld de waarde $b = +1$, men kan dan deze fractie schrijven als:

$$f_{J_o}(\alpha) = \frac{\int_{J_o}^{+\infty} dj P(j, +1)}{\int_{J_o}^{+\infty} dj P(j)}, \quad (3.8)$$

waarbij $P(j) dj = dj \int_{-\infty}^{+\infty} db P(j, b)$. Hier is expliciet aangeduid dat f_{J_o} afhankelijk is van α via de ordeparameters in de uitdrukking voor $P(j, b)$. Bemerkt dat als $J_o = 0$ de grootheid $f_{J_o=0}$ gelijk is aan de fractie van componenten die correct de overeenkomstige component van de MSB voorspellen bij volledig knippen van de precursor. Het resultaat voor f_{J_o} is gegeven door (Bijlage B):

$$f_{J_o}(\alpha) = \frac{1}{2} + \frac{1}{2H(J_o)} \int_{-\infty}^{+\infty} D\psi \tanh(\sqrt{\hat{Q}}\psi) H\left(\frac{J_o - \hat{\gamma}\psi}{\sqrt{1 - \hat{\gamma}^2}}\right). \quad (3.9)$$

3. Leren op de hypersfeer

Om nu de kwaliteitsmaat f_x te berekenen, volstaat het J_o te bepalen uit $x = 2 \int_{J_o}^{+\infty} P(j) dj = 2 H(J_o)$ en deze in te vullen in de vorige uitdrukking (3.9). Dit verzekert dat er enkel een fractie x geknipt wordt. Bij volledig knippen is $x = 1$. De waarde van J_o volgt dan uit $H(J_o) = 1/2$. Zodat $J_o = 0$. Er geldt dus:

$$f_{x=1}(\alpha) = f_{J_o=0} = \frac{1}{2} + \int_{-\infty}^{+\infty} D\psi \tanh(\sqrt{\hat{Q}}\psi) H\left(\frac{-\hat{\gamma}\psi}{\sqrt{1-\hat{\gamma}^2}}\right). \quad (3.10)$$

De uitdrukkingen (3.9) en (3.10) zijn de uiteindelijk gezochte uitdrukkingen voor de kwaliteitsmaat $f_x(\alpha)$, respectievelijk voor $x < 1$ (3.9) en $x = 1$ (3.10). Zij zullen straks gebruikt worden om verschillende precursors op de hypersfeer met mekaar te vergelijken. De afhankelijkheid van f_x van de potentiaal $V(\lambda)$ zit volledig in de parameter $\hat{\gamma}$.

Het is tevens mogelijk een bovengrens af te leiden voor f_x . De waarde van de integraal in (3.9) neemt toe² met $\hat{\gamma}$. Hieruit volgt onmiddellijk dat men dus een bovengrens kan afleiden voor de fractie f_x door de limiet $\hat{\gamma} \rightarrow 1$ te nemen:

$$f_{J_o}(\alpha) \leq \frac{1}{2} + \frac{1}{2H(J_o)} \int_{J_o}^{+\infty} D\psi \tanh(\sqrt{\hat{Q}}\psi), \quad (3.11)$$

en tevens voor volledig knippen:

$$f_{x=1}(\alpha) \leq \frac{1}{2} + \int_0^{+\infty} D\psi \tanh(\sqrt{\hat{Q}}\psi). \quad (3.12)$$

Deze bovengrenzen hangen enkel af van de ordeparameter $\hat{Q}$ van de MSB, en kunnen dus niet overschreden worden, wat ook de keuze voor de potentiaal $V(\lambda)$ is.

²Beschouw $J_o = 0$. Voor $J_o \neq 0$ is de redenering analoog. Het integrandum van de integraal is gegeven door $\tanh(\sqrt{\hat{Q}}\psi) H\left(\frac{-\hat{\gamma}\psi}{\sqrt{1-\hat{\gamma}^2}}\right) \cdot \exp\left(-\frac{\psi^2}{2}\right)$. Merk vooreerst op dat $0 \leq H(x) \leq 1$. Verder geldt er dat $H(x) \rightarrow 1$ als $x \rightarrow -\infty$ en $H(x) \rightarrow 0$ als $x \rightarrow +\infty$ en $H(0) = 1/2$. De enige term die afhangt van $\hat{\gamma}$ is $H\left(\frac{-\hat{\gamma}\psi}{\sqrt{1-\hat{\gamma}^2}}\right)$. Merk op dat de H -functie zijn waarde van $1/2$ bereikt bij $\psi = 0$. Als $\hat{\gamma}$ naar zijn maximale 1 gaat, nadert $H\left(\frac{-\hat{\gamma}\psi}{\sqrt{1-\hat{\gamma}^2}}\right)$ naar $\Theta(\psi)$. De bijdrage tot de integraal van het interval $\psi > 0$, waar de tanh-functie positief is wordt dus steeds groter met toenemende $\hat{\gamma}$. Anderzijds wordt de bijdrage tot de integraal van het interval $\psi < 0$, waar de tanh-functie negatief is steeds kleiner.

3.3 De MSN als precursor

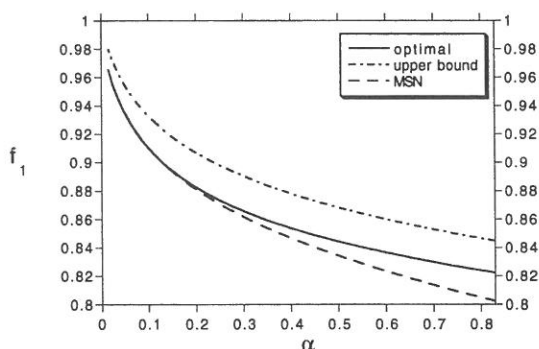
Laten we nu de hierboven bekomen resultaten eens toepassen op de MSN. Kies dus de volgende leerregel:

$$\begin{aligned} V(\lambda) &= (\kappa_s - \lambda)^2 \Theta(\kappa_s - \lambda) \\ P(\mathbf{J}) &= \delta(\mathbf{J}^2 - N), \end{aligned} \quad (3.13)$$

hierbij is κ_s genoteerd om aan te duiden dat het de stabiliteit betreft van de sferische koppelingen, in tegenstelling tot κ_b , de stabiliteit van de binaire koppelingen. Bemerk wel dat de kostfunctie van (2.3) met deze potentiaal over het algemeen geen uniek minimum heeft op de hypersfeer. Inderdaad voor $\alpha < \alpha_c(\kappa_s)$ zijn er meerdere vektoren die de kostfunctie minimaliseren [10]. Het is pas bij $\alpha = \alpha_c(\kappa_s)$ dat er één unieke vektor is die de kostfunctie minimaliseert. Deze vektor is de MSN, de gewenste precursor. De reeds genomen limiet $\beta \rightarrow \infty$ en $q \rightarrow 1$ is dus enkel geldig indien $\alpha = \alpha_c(\kappa_s)$, maar deze voorwaarde is niet noodzakelijk voldaan bij het nemen van de limiet. Gebruik maken van de bovenstaande leerregel (3.13) in de limiet $\beta \rightarrow \infty$ en $q \rightarrow 1$, impliceert dus niet onmiddellijk het vinden van de MSN. We moeten net zoals bij de MSB een voorwaarde opleggen om te verzekeren een oplossing te vinden bij verzadiging van het netwerk. In dit geval hebben Gardner en Derrida getoond dat het volstaat de limiet voor $x \rightarrow \infty$ te nemen in formule (3.3) [10].

Men kan nu de kwaliteitsmaat $f_x(\alpha)$ uit vorige sectie berekenen. Figuur 3.1 toont de kwaliteitsmaat $f_x(\alpha)$ voor $x = 1$, bij volledig knippen van de MSN. De bovengrens voor $\hat{\gamma} = 1$ is tevens toegevoegd. Deze resultaten voor het volledig knippen van de MSN zijn reeds bekomen door Penney en Sherrington en waren reeds vermeld in sectie 2.1. Bij $\alpha = 0.83$ zijn er 80% van de componenten correct voorspeld. Dit percentage neemt toe tot ongeveer 90% bij $\alpha = 0.1$. Voor nog kleinere α -waarden wordt de bovengrens dichter en dichter benaderd. Figuur 3.2 toont de fractie van correct geknipte componenten als enkel de xN grootste componenten van de MSN geknipt worden. Curves zijn gegeven voor $x = 0.4$ (bovenste curves), 0.6 (middenste curves) en 0.8 (onderste curves). (Dit is dus voor het knippen van 40, 60 en 80% van de grootste componenten.) De bovengrens voor $\hat{\gamma} \rightarrow 1$ is weer gegeven om te vergelijken. Figuur 3.2 toont duidelijk dat de grootste componenten van de MSN een zeer grote waarschijnlijkheid hebben om correct de overeenkomstige component van de MSB te voorspellen. Des te kleiner de fractie van grootste

3. Leren op de hypersfeer



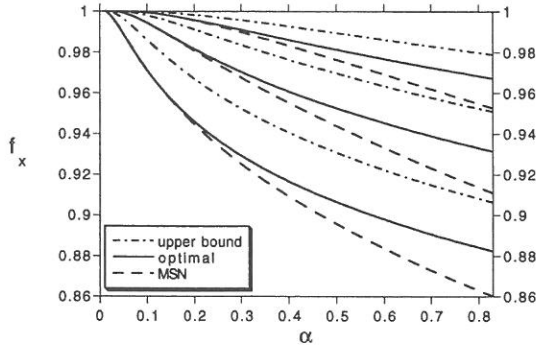
Figuur 3.1: De fractie $f_1(\alpha)$ van correct voorspelde componenten bij volledig knippen van de hypersfeer precursor. Resultaten zijn getoond voor de MSN en de optimale precursor bepaald door (3.16). De bovengrens voor $\hat{\gamma} \rightarrow 1$ (3.12) is ook gegeven.

componenten die men knipt des te groter het aantal correct voorspelde componenten in deze fractie. Zo geldt bijvoorbeeld dat er 91% van de synapsen van de MSB correct voorspeld worden als men de 60% grootste componenten van de MSN knipt. Dit aantal stijgt tot ongeveer 95% als men slechts de 40% grootste aan een knip-procedure onderwerpt. De aanname van Penny en Sherrington [26] dat de grootste componenten van de MSN bij knippen, de grootste waarschijnlijkheid hebben om correct de overeenkomstige component van de MSB te voorspellen wordt dus duidelijk bevestigd in figuur 3.2. Wat zij concludeerden uit numerieke simulaties voor kleine systemen $N \leq 25$ blijft geldig in de limiet voor $N \rightarrow \infty$.

3.4 De optimale precursor op de hypersfeer

We wensen nu de optimale potentiaal te bepalen die de grootst mogelijke overlap oplevert met de MSB. We gaan daarvoor terug naar de zadelpuntvergelijkingen (3.2) en (3.4) voor een algemene potentiaal. Combineren van deze twee vergelijkingen levert [17]:

3.4. De Optimale Precursor



Figuur 3.2: De fractie $f_x(\alpha)$ van correct voorspelde componenten als enkel de xN grootste componenten van de hypersfeer precursor geknipt worden. Resultaten zijn getoond voor de MSN en de optimale precursor bepaald door (3.16) bij (van boven naar onder) $x = 0.4, 0.6$ en 0.8 . De bovengrens voor $\hat{\gamma} \rightarrow 1$ (3.11) is ook gegeven.

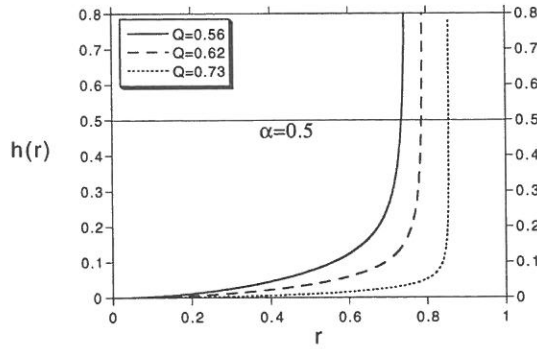
$$r^2 = \alpha \frac{\left[\int_{-\infty}^{+\infty} D\sigma [\lambda_o - \sigma] \cdot g_r(\sigma) \right]^2}{\int_{-\infty}^{+\infty} D\sigma (\lambda_o(\sigma, x) - \sigma)^2}. \quad (3.14)$$

Gebruik makend van de Schwartz ongelijkheid (E.26) volgt er dan:

$$r^2 \leq \alpha \int_{-\infty}^{+\infty} D\sigma (g_r(\sigma))^2. \quad (3.15)$$

De gelijkheid geldt als de potentiaal $V(\lambda)$ zodanig gekozen is dat $\lambda_o - \sigma = C g_r(\sigma)$ met C een constante. Bemerkt echter dat de uitdrukking (3.15) onafhankelijk is van die gekozen precursor potentiaal. Indien het rechterlid van (3.15) onafhankelijk zou zijn van r geldt er onmiddellijk dat de gelijkheid in (3.15) een bovengrens is voor r . Jammer genoeg is het rechterlid wel afhankelijk van r . Kunnen we alsnog bewijzen dat de gelijkheid een bovengrens is voor r ? Het antwoord is ja. Figuur 3.3 toont de functie $h(r) \equiv r^2 / \int D\sigma (g_r(\sigma))^2$. Het is een strikt monotoon stijgende functie van r . In figuur 3.3 is tevens een

3. Leren op de hypersfeer



Figuur 3.3: De functie $h(r) \equiv r^2 / \int D\sigma (g_r(\sigma))^2$ voor verschillende waarden van Q . De horizontale lijn representeert $\alpha = 0.5$ op de y -as zoals beschreven in de tekst.

horizontale lijn getekend die $\alpha = 0.5$ representeert op de y -as. (De α waarde is volledig arbitrair gekozen.) Vergelijking (3.15) impliceert dat de mogelijke r waarden beperkt zijn tot waarden waarvoor $h(r) \leq \alpha$. Op de figuur zijn dus alle r waarden die links liggen van het snijpunt van de horizontale lijn met de functie $h(r)$ mogelijke r waarden. (voor $\alpha = 0.5$) Vermits $h(r)$ nu monotoon stijgend is geldt er dus dat de gelijkheid $h(r) = \alpha$ optreed bij een maximale waarde van r . De gelijkheid bepaalt door de potentiaal $V(\lambda)$ waarvoor dat $\lambda_o - \sigma = C g_r(\sigma)$, is dus een bovengrens voor r . Merk op dat indien $h(r)$ monotoon dalend is of helemaal niet monotoon is, een dergelijke uitspraak niet kan gedaan worden.

De maximaal mogelijke overlap r die dus kan bekomen worden binnen de beschouwde klasse van algoritmen is oplossing van de vergelijking:

$$r^2 = \alpha \int_{-\infty}^{+\infty} D\sigma (g_r(\sigma))^2. \quad (3.16)$$

Vermits deze uitdrukking onafhankelijk is van de potentiaal kunnen we r_{max} berekenen zonder expliciete kennis van de optimale potentiaal.

Figuur 3.1 en 3.2 tonen de kwaliteitsmaat f_x respectievelijk voor $x = 1$ en voor $x = 0.4, 0.6$ en 0.8 als functie van α voor de optimale precursor bepaald door (3.16). De bovengrens bepaald door $\hat{\gamma} = 1$ is tevens toegevoegd. Uit figuur

3.4. De Optimale Precursor

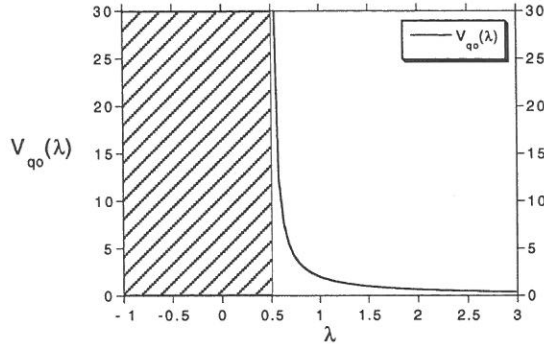
3.2 volgt onmiddellijk dat de aanname van Penney en Sherrington voor de MSN ook geldig is voor de optimale precursor: de grootste componenten hebben de grootste waarschijnlijkheid om bij knippen correct de overeenkomstige component van de MSB te voorspellen. Het verschil tussen deze precursor en de MSN wordt enkel zichtbaar bij grote waarden van α . Nabij de maximale geheugen-capaciteit $\alpha_c = 0.83$ van het binair perceptron is het verschil ongeveer 2%. Dit impliceert dus dat deze optimale precursor een groter aantal componenten van de MSB correct voorspelt in vergelijking met de MSN. Dit geldt zowel voor volledig knippen van de precursor als bij het knippen van een fractie van de grootste componenten.

3.4.1 Een quasi-optimale potentiaal

De kwaliteitsmaat bekomen voor de maximaal mogelijke overlap gegeven door (3.16) is niet van praktisch nut voor simulaties vermits we de expliciete vorm van de potentiaal $V(\lambda)$ niet kennen. Het is evenwel mogelijk deze te verkrijgen met variationele methodes [17]. In plaats van deze weg te bewandelen kan men trachten een uitdrukking voor deze potentiaal voorop te "raden". Beschouw even het Gibbs ensemble van continue vektoren $\mathbf{J}$ die de patronen stabiliseren met een minimale stabiliteit κ_b . Uit figuur 1.3 volgt dat $\kappa_b < \kappa_s$ zodat er een heel gebied van continue vektoren is dat de patronen stockeert met stabiliteit κ_b . In dit Gibbs ensemble zitten natuurlijk ook de binaire vektoren die we zoeken. Als deze binaire vektoren homogeen verdeeld zouden zijn over het Gibbs ensemble dan zou het zwaartepunt ervan samenvallen met het zwaartepunt van het gehele ensemble. Dit zwaartepunt zou dus een maximale overlap met de vektor $\langle \mathbf{B} \rangle$ opleveren. Nu blijkt echter dat de ordeparameter $q < Q$, en dit voor alle waarden van α . Hieruit volgt dat de binaire vektoren slechts over een exponentieel klein gebied van het totale Gibbs ensemble verdeeld zijn. Hoedanook het zwaartepunt van het Gibbs ensemble is de beste keuze van precursor die men kan maken zonder verdere informatie over de ligging van het gezochte binair ensemble [32]. Men verwacht dan ook dat de optimale potentiaal een precursor moet opleveren dicht in de buurt van dit zwaartepunt. Zoals in het geval van leren met een leraar [4] kan men dan trachten dit zwaartepunt te vinden met een volgende potentiaal³:

³In bijlage D wordt getoond dat deze potentiaal convex is op de hypersfeer. Hieruit volgt dan dat het minimum van de potentiaal uniek is. De aanname dat voor $\beta \rightarrow \infty, q \rightarrow f = 1$ is dus geldig.

3. Leren op de hypersfeer



Figuur 3.4: De quasi-optimale potentiaal $V_{qo}(\lambda)$ op de hypersfeer voor $\kappa_b = 0.5$.

$$V_{qo}(\lambda) \equiv \begin{cases} \frac{1}{\lambda - \kappa_b} & : \text{als } \lambda \geq \kappa_b \\ +\infty & : \text{als } \lambda < \kappa_b. \end{cases} \quad (3.17)$$

Deze potentiaal is getoond in figuur 3.4 voor $\kappa_b = 0.5$. Vermits de randen van het Gibbs ensemble gevormd wordt door die vektoren waarvoor $\text{Min}_\mu \lambda_\mu = \kappa_b$ "duwt" deze potentiaal de precursor vektor van de randen weg naar het zwaartepunt.

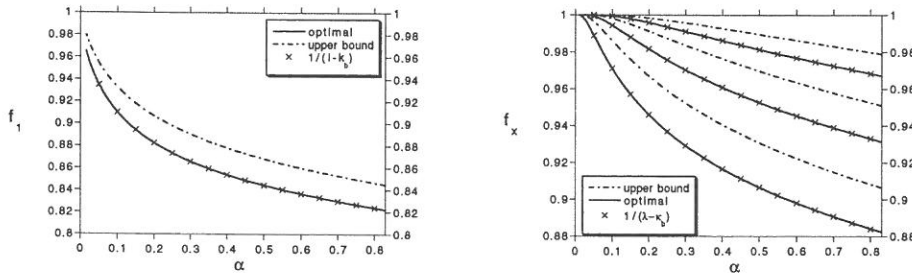
Figuur 3.5 toont de fractie f_1 en f_x (voor $x = 0.4, 0.6$ en 0.8) voor de optimale potentiaal bekomen uit de maximale overlap (3.16) (zoals in figuur 3.1) tesamen met de resulaten bekomen voor (3.17). De kurven vallen zo goed als samen. We zullen de potentiaal (3.17) de quasi-optimale potentiaal $V_{qo}(\lambda)$ noemen⁴.

3.5 Simulaties

Tot hiertoe hebben we ons geconcentreerd op het *aantal* binaire synapsen dat we correct kunnen voorspellen door het knippen van een continue precursor.

⁴De subindex "qo" staat voor: quasi-optimaal.

3.5. Simulaties



Figuur 3.5: De fractie $f_1(\alpha)$ en $f_x(\alpha)$ (voor $x = 0.4, 0.6$ en 0.8) van correct voorspelde componenten van de hypersfeer precursor. Resultaten zijn getoond voor de optimale precursor bepaald door (3.16) en zijn benadering gegeven door (3.17)

Dit is natuurlijk een interessante kwaliteitsmaat maar voor praktische doeleinden is het echter veel interessanter van zich af te vragen wat de maximaal mogelijke *stabiliteit* is die nog bekomen kan worden na het knippen van een fractie van de grootste componenten. We zullen dit onderzoeken met numerieke simulaties. Het is duidelijk dat het berekenen van de nog mogelijke maximale stabiliteit na het knippen van een deel van de synapsen geen enkele benadering meer toelaat in het bepalen van de overige componenten. Men moet dus een enumeratie doen op deze resterende componenten terwijl de geknipte synapsen vast gehouden worden. De enumeratie methode die we hier gebruiken is het Gray-code algoritme [30]. De numerieke simulaties zijn als volgt uitgevoerd.

1. Kies de grootte van het netwerk, N . (Hier $N=50$ gekozen.)
2. Kies de waarde van $\alpha = p/N$.
3. Kies een set van $p = \alpha N$ patronen η^μ . De simulaties zijn uitgevoerd met Gaussische patronen om finite size effecten te verkleinen [22].
4. Bepaal vervolgens de gewenste precursor. Voor het bepalen van de MSN werd het AdaTron algoritme [2] gebruikt. In het geval van de quasi-optimale precursor werd gebruik gemaakt van een standaard gradient-descent algoritme. Deze minimalisatie van de kostfunctie is snel en eenvoudig omdat de kostfunctie een uniek minimum heeft op de hypersfeer.

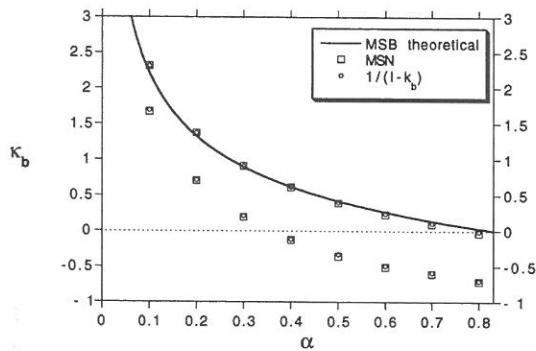
3. Leren op de hypersfeer

5. Zodra de precursor berekend is kunnen we ze zijn componenten in absolute waarde rangschikken van groot naar klein. Dit zal straks het knip proces vereenvoudigen.
6. Het is noodzakelijk van dezelfde herrangschikking van de componenten uit te voeren op de patronen $\boldsymbol{\eta}^\mu$ zodat de scalaire produkten $\mathbf{J} \cdot \boldsymbol{\eta}^\mu$ en dus de stabiliteiten dezelfde waarde blijven behouden.
7. Knip een zeker percentage van de grootste componenten. In deze simulaties zijn er steeds 60% van de componenten geknipt. Dit wil dus zeggen dat men de *eerste* 30 componenten ($N = 50$) van de precursor moet knippen. (Denk aan de herrangschikking van groot naar klein die we daarstraks uitgevoerd hebben.) Deze componenten worden tijdens de overige stappen in het algoritme vast gehouden, hierbij veronderstellend dat deze componenten betrouwbare voorspellers zijn van de overeenkomstige MSB synapsen. Het hoofddoel van de simulaties is van dit laatste te verifiëren.
8. De overige niet geknipte componenten worden vrij gehouden om deel te nemen aan een volledige enumeratie met het "Gray-code" algoritme. Vermits er 30 componenten geknipt zijn, wordt deze enumeratie uitgevoerd op de 20 andere en kleinste componenten. Tijdens de enumeratie wordt de binaire vektor met de grootste minimale stabiliteit κ_b geselecteerd als benadering voor de MSB.
9. Ga terug naar stap 3 en herhaal de volgende stappen in het algoritme. Doe dit 100 maal.
10. Bereken het gemiddelde van de 100 bekomen κ_b waarden.

Dit algoritme dient herhaald te worden voor verschillende waarden van α . De gevonden gemiddelde waarden van κ_b worden dan uitgezet als functie van de parameter α . Het resultaat is gegeven in figuur 3.6.

Figuur 3.6 toont de resultaten voor de MSN en de quasi-optimale potentiaal. Naast de strategie van "knippen en enumereren" zijn er ook resultaten getoond voor het volledig knippen van de beide precursors (de onderste datapunten op figuur 3.6). In dit geval wordt er na het bekomen van de precursor (stap 4) een volledig knip-proces uitgevoerd op de componenten waarnaar rechtstreeks overgegaan wordt naar de berekening van de minimale stabiliteit. (Dus stappen 5 tot en met 8 worden dan overgeslagen.) Zoals verwacht geeft volledig knippen

3.5. Simulaties



Figuur 3.6: Minimale stabiliteit κ_b als een functie van α bepaald door de numerieke simulaties beschreven in de tekst voor een perceptron met $N = 50$. De onderste datapunten geven de resultaten bij volledig knippen. Er is telkens gemiddeld over 100 setjes van patronen.

een slechte ondergrens voor κ_b . Er is een opmerkelijke verbetering met de strategie van "knippen en enumereren". Er is bijna volledige overeenstemming met de theoretische curve. Voor kleine waarde van α is dit zeker waar, maar de overeenstemming vermindert bij grotere waarde van α zoals natuurlijk verwacht is. Immers er is een steeds groter wordende fractie van componenten die fout geknipt zijn.

Opmerkelijk is het zeer kleine verschil tussen de MSN en de quasi-optimale precursor. De laatste heeft een heel kleine verbetering op de bekomen minimale stabiliteit. Dit is niet verwonderlijk omdat slechts 30 (60% van $N = 50$) componenten geknipt werden. Hiervan zal de optimale precursor er maximaal 2% meer correct voorspellen. (Dit is slechts 0.6 componenten.) We verwachten dat het verschil wel groter wordt met toenemende waarde van N .

3. Leren op de hypersfeer

Hoofdstuk 4

Leren in de hyperkubus

4.1 Inleiding

In dit hoofdstuk zullen we trachten nog betere precursors voor de MSB te bekomen. Een verbetering binnen de beschouwde klasse van kostfuncties (2.3) op de hypersfeer is onmogelijk vermits in vorig hoofdstuk de optimale precursor op deze constraint bepaald is. Om toch nog een betere precursor te bekomen kan men bijvoorbeeld de vorm van de kostfunctie $E(\mathbf{J})$ aanpassen. Men beschouwt dan een andere, of meer uitgebreide klasse, van kostfuncties. Pérez Vicente, Carrabina and Valderrama [27] gebruikten een gewijzigde kostfunctie die de term $\sum_i (J_i^2 - 1)^2$ bevat. Deze verschuift het minimum dichter naar de binaire vektoren toe. Deze extra term veroorzaakt echter een groot aantal lokale minima waarin de minimalisatie kan vast geraken.

Een ander alternatief is van de beschouwde klasse van kostfuncties ongewijzigd te laten maar van het minimalisatiedomein te veranderen. Dit is de aanpak die ik zal volgen. In dit hoofdstuk zal ik de hypersfeer $\mathbf{J}^2 = N$ als minimalisatiedomein vervangen door de hyperkubus gegeven door:

$$|J_i| \leq 1 \quad \forall i : 1 \dots N. \quad (4.1)$$

Het minimalisatiedomein verandert dus van het oppervlak van de hypersfeer naar het volume van de ingeschreven hyperkubus. Er zijn twee eigenschappen die de hyperkubus tot een aantrekkelijke keuze voor het minimalisatiedomein maken.

In het geval van de hypersfeer is alle informatie over de richting van de binaire vektoren verloren gegaan, maar bij de hyperkubus is dit niet het geval.

4. Leren in de hyperkubus



Figuur 4.1: Terwijl er op de hypersfeer geen onderscheid gemaakt wordt tussen de binaire vectoren B en de continue vectoren J is dit wel het geval op de hyperkubus. Hier zijn de binaire vectoren bepaald door de hoekpunten van de hyperkubus.

De binaire vectoren zijn die richtingen die wijzen naar de 2^N hoekpunten van de N -dimensionale hyperkubus (zie figuur 4.1). Bovendien zijn deze binaire vectoren ook nog de langste uit het hele minimalisatiedomein, zodat ze een voordeel hebben in het genereren van grotere stabiliteiten λ_μ . (Bemerkt dat de stabiliteiten in de hyperkubus gedefinieerd blijven als $\lambda_\mu = \frac{1}{\sqrt{N}} \mathbf{J} \cdot \boldsymbol{\eta}^\mu$ en dus niet genormeerd zijn met de lengte van de vektor $\mathbf{J}$ in de hyperkubus. Op deze manier hebben dus langere vectoren inderdaad grotere stabiliteiten.) Dus bij elke kostfunctie die kleine stabiliteiten discrimineert zullen de binaire vectoren als hoekpunten van de hyperkubus een competitief voordeel hebben op de andere vectoren. Het is dus redelijk van te verwachten dat het minimum van een dergelijke kostfunctie als precursor een grotere overlap heeft met de MSB.

De tweede interessante eigenschap van de hyperkubus is dat het een convex minimalisatiedomein is. Zodat elke kostfunctie van de vorm $E(\mathbf{J}) = \sum_\mu V(\lambda_\mu)$ met een convexe potentiaal een uniek minimum heeft dat eenvoudig met een standaard gradient descent algoritme kan bepaald worden.

Beschouw een potentiaal $V(\lambda)$ met een uniek minimum op de hyperkubus. De distributie (4.1) die het minimalisatiedomein definieert is in dit geval een lokale constraint zodat de opmerking van sectie 2.4.4 niet van toepassing is.

4.2 Het $V(\lambda)$ /MSB systeem

Ik zal hier de zadelpuntvergelijkingen voor het systeem van de MSB en de willekeurige potentiaal, alsook de uiteindelijk gezochte kwaliteitsmaat $f_x(\alpha)$, beschrijven. Voor details betreffende de berekeningen verwijs ik naar Bijlage C.

De zadelpuntvergelijkingen voor de twee systemen afzonderlijk

De zadelpuntvergelijkingen voor het precursor systeem in de hyperkubus komen uit (2.9–2.12) met $\beta \rightarrow \infty$. Dan $q \rightarrow f$ en $\hat{q}$ samen met $2\hat{f} + \hat{q}$ gaan naar oneindig. Daarom zijn er drie nieuwe parameters ingevoerd om q , $\hat{f}$, en $\hat{q}$ te vervangen:

$$\begin{aligned} x &= \beta(f - q), \\ y &= (2\hat{f} + \hat{q})(f - q), \\ t &= \frac{\sqrt{\hat{q}}}{2\hat{f} + \hat{q}}. \end{aligned} \quad (4.2)$$

Men bekomt aldus de volgende 4 zadelpuntvergelijkingen voor x, f, t , en y :

$$y = -\alpha \int_{-\infty}^{+\infty} D\sigma \left[\frac{\partial \lambda_o}{\partial \sqrt{f}\sigma} (\sqrt{f}\sigma, x) - 1 \right], \quad (4.3)$$

$$(ty)^2 = \alpha \int_{-\infty}^{+\infty} D\sigma \left[\lambda_o(\sqrt{f}\sigma, x) - \sqrt{f}\sigma \right]^2, \quad (4.4)$$

$$y = 1 - 2H\left(\frac{1}{t}\right) = \text{Erf}\left(\frac{1}{t\sqrt{2}}\right), \quad (4.5)$$

$$f = 1 - y + yt^2 - \sqrt{\frac{2}{\pi}} t \exp\left[\frac{-1}{2t^2}\right]. \quad (4.6)$$

Hierin is λ_o gegeven door:

$$\lambda_o(\sqrt{f}\sigma, x) = \text{Arg Min}_{\lambda} \left[V(\lambda) + \frac{(\lambda - \sqrt{f}\sigma)^2}{2x} \right]. \quad (4.7)$$

In vergelijking met de minimalisatie van de kostfunctie $E(\mathbf{J})$ op de hypersfeer waar enkel de parameter x voorkomt, komen hier dus 3 extra ordeparameters

4. Leren in de hyperkubus

naar voren. De betekenis van de parameter f is duidelijk en volgt onmiddellijk uit (2.8): het is de zelfoverlap van de precursor. Wat de betekenis van de ordeparameters t en y is wordt duidelijk als we kijken naar de distributie $P(j)$ van de precursor componenten in de hyperkubus:

$$P(j) = \Theta(1 - j^2) \frac{1}{\sqrt{2\pi t}} e^{\frac{-j^2}{2t^2}} + H(1/t)(\delta(j - 1) + \delta(j + 1)). \quad (4.8)$$

De componenten van de precursor die in absolute waarde kleiner zijn dan 1 zijn Gaussisch verdeeld met een variantie t^2 . De twee staarten van de Gaussische verdeling zijn samengepakt in twee Dirac-delta pieken bij $+1$ en -1 . Terugkijkend naar de zadelpuntvergelijking (4.5) volgt dan dat $1 - y$ de fractie van componenten van $\mathbf{J}$ vertegenwoordigt die in absolute waarde gelijk zijn aan 1. Bemerk dat de vorm van de distributie $P(j)$ varieert met α via zijn afhankelijkheid van de ordeparameter t . Hieruit volgt dat de fractie van componenten groter dan een vaste waarde $J_o > 0$, gegeven door $2 \int_{J_o}^{+\infty} P(j) dj = 2H(J_o/t)$, ook varieert met α . Ik kom hier later nog op terug.

De vergelijkingen voor het MSB systeem blijven natuurlijk bepaald door (2.14-2.16).

De zadelpuntvergelijkingen voor r en $\hat{r}$

De zadelpuntvergelijking voor r wordt gegeven door:

$$r = \int_{-\infty}^{+\infty} D_{\hat{\gamma}}(\varphi, \psi) \cdot \tanh(\sqrt{\hat{Q}}\psi) \cdot J_{\min}(t\varphi), \quad (4.9)$$

en deze voor $\hat{r}$ (of $\hat{\gamma} = \hat{r}/\sqrt{\hat{q}\hat{Q}}$):

$$\hat{\gamma}\sqrt{\hat{Q}}(1 - Q)ty = \alpha \int_{-\infty}^{+\infty} D\sigma \left[\lambda_o(\sqrt{f}\sigma, x) - \sqrt{f}\sigma \right] g_r(\sigma). \quad (4.10)$$

Hierbij is $g_r(\sigma)$ zoals in (3.5) en $J_{\min}$ gegeven door:

$$J_{\min}(t\varphi) = \text{Arg} \min_{J \in [-1, 1]} \left[\frac{1}{2}J^2 - tJ\varphi \right] = \begin{cases} t\varphi & : \text{als } |t\varphi| \leq 1 \\ +1 & : \text{als } t\varphi > 1 \\ -1 & : \text{als } t\varphi < -1. \end{cases} \quad (4.11)$$

4.2. Het $V(\lambda)$ /MSB systeem

De vergelijking (4.9) hangt niet expliciet af van de precursor potentiaal $V(\lambda)$. Vergelijking (4.10) echter wel en dit via de functie $\lambda_o(\sqrt{f}\sigma, x)$ bepaald door (4.7). In Bijlage C, sectie C.2.1 is getoond dat de zadelpuntvergelijking (4.9) uitdrukt dat r de gemiddelde waarde is van het produkt van de overeenkomstige componenten van $\mathbf{J}$ en $\mathbf{B}$: $\langle bj \rangle = \sum_b \int dj P(j, b) bj$. (Dit is zoals getoond in dezelfde sectie ook waar voor de zadelpuntvergelijking voor r op de hypersfeer.)

Zoals reeds vermeld in hoofdstuk 2 en 3 moeten dus eerst weer de ordeparameters van beide systemen (precursor en MSB) bepaald worden, als functie van α , uit hun respectievelijke zadelpuntvergelijkingen. Vervolgens dienen r en $\hat{r}$ berekend te worden als functie van α . Zodra de volledige set van ordeparameters gekend is kan men overgaan tot het berekenen van de kwaliteitsmaat f_x m. b. v. de gevonden uitdrukking voor $P(j, b)$.

Uitdrukking voor de kwaliteitsmaat f_x

De uitdrukking voor $P(j, b)$ is bekomen in sectie C.3 en is gegeven door:

$$P(j, b) = \int_{-\infty}^{+\infty} D_{\hat{\gamma}}(\varphi, \psi) \cdot \Delta(j, t\varphi) \frac{1}{2} \left[1 + \tanh(\sqrt{\hat{Q}} b \psi) \right], \quad (4.12)$$

waarbij:

$$\Delta(j, t\varphi) = \begin{cases} \delta(j - t\varphi) & : \text{als } |t\varphi| \leq 1 \\ \delta(j - 1) & : \text{als } t\varphi > 1 \\ \delta(j + 1) & : \text{als } t\varphi < -1. \end{cases} \quad (4.13)$$

Met deze uitdrukking kan men nu weerom de fractie van componenten $|J| > J_o > 0$ berekenen die bij knippen correct de corresponderende component van de MSB opleveren. Vermits $P(j, b) = P(-j, -b)$ moeten we enkel rekening houden met bijvoorbeeld de waarde $b = +1$. Men kan dan deze fractie weer schrijven zoals (3.8):

$$f_{J_o}(\alpha) = \frac{\int_{J_o}^{+\infty} dj P(j, +1)}{\int_{J_o}^{+\infty} dj P(j)}, \quad (4.14)$$

waarbij $P(j)$ gegeven is door (4.8). Hier is weer expliciet aangeduid dat f_{J_o} afhankelijk is van α via de ordeparameters in de uitdrukking voor $P(j, b)$. Merk dat als $J_o = 0$ de grootheid $f_{J_o=0}$ gelijk is aan de fractie van componenten die correct de overeenkomstige component van de MSB voorspellen bij volledig knippen van de precursor. Het resultaat voor f_{J_o} is gegeven door (bijlage C):

4. Leren in de hyperkubus

$$f_{J_o}(\alpha) = \frac{1}{2} + \frac{1}{2H(J_o/t)} \int_{-\infty}^{+\infty} D\psi \tanh(\sqrt{\hat{Q}}\psi) H\left(\frac{J_o/t - \hat{\gamma}\psi}{\sqrt{1 - \hat{\gamma}^2}}\right), \quad (4.15)$$

en bij volledig knippen ($J_o = 0$) geldt er dan:

$$f_{x=1}(\alpha) = \frac{1}{2} + \int_{-\infty}^{+\infty} D\psi \tanh(\sqrt{\hat{Q}}\psi) H\left(\frac{-\hat{\gamma}\psi}{\sqrt{1 - \hat{\gamma}^2}}\right). \quad (4.16)$$

Merk op dat deze laatste uitdrukking dezelfde is als (3.10) voor volledig knippen op de hypersfeer. De uitdrukkingen (4.15) en (4.16) zijn de uiteindelijk gezochte uitdrukkingen voor de kwaliteitsmaat $f_x(\alpha)$, respectievelijk voor $x < 1$ (4.15) en $x = 1$ (4.16). Zij zullen straks gebruikt worden om verschillende precursors in de hyperkubus met mekaar te vergelijken. De afhankelijkheid van f_{J_o} van de potentiaal $V(\lambda)$ zit volledig in de parameter $\hat{\gamma}$ en t . Om nu de kwaliteitsmaat f_x te berekenen, volstaat het J_o te bepalen uit $x = 2 \int_{J_o}^{+\infty} P(j) dj = 2H(J_o/t)$. Deze gevonden waarde voor J_o invullen in de vorige uitdrukking (4.15) verzekert dat er enkel een fractie x geknipt wordt. Bemerkt echter dat voor een vaste waarde van J_o deze fractie x verandert met α omdat t verandert met α . Dus als we de grootheid $f_x(\alpha)$ wensen te bekomen moeten we niet J_o vast houden maar de verhouding J_o/t . Enkel dan is een constante fractie van x geknipte componenten gewaarborgd als functie van α . Als straks specifieke keuzes voor de potentiaal vooropgesteld worden zal ik hier mee rekening houden tijdens het berekenen van de kwaliteitsmaat.

Men kan nu tevens ook weer een bovengrens afleiden voor de kwaliteitsmaat. Analoog zoals in hoofdstuk 3 geldt hier ook dat de waarde van de integraal in (4.15) toeneemt met $\hat{\gamma}$. Hieruit volgt onmiddellijk dat een bovengrens voor (4.15) gegeven wordt door de limiet $\hat{\gamma} \rightarrow 1$ te nemen:

$$f_{J_o}(\alpha) \leq \frac{1}{2} + \frac{1}{2H(J_o/t)} \int_{J_o/t}^{+\infty} D\psi \tanh(\sqrt{\hat{Q}}\psi), \quad (4.17)$$

en in het geval van volledig knippen geldt er weer:

$$f_{x=1}(\alpha) \leq \frac{1}{2} + \int_0^{+\infty} D\psi \tanh(\sqrt{\hat{Q}}\psi). \quad (4.18)$$

Vermits de uitdrukking voor $f_{x=1}$ (4.16) dezelfde is als (3.10) op de hypersfeer geldt er dat de bovengrens (4.18) geldig is voor gelijk welke potentiaal op de hypersfeer en de hyperkubus.

4.3 De MSC als Precursor

De eerste suggestie voor een precursor op de hypersfeer in vorig hoofdstuk was het maximaal stabiel perceptron, de MSN. Men kan op eenvoudige wijze het idee van het maximaal stabiel perceptron op de hyperkubus invoeren. Het is dan gedefinieerd als de vektor $\mathbf{J}$ met $|J_i| \leq 1 \forall i$ die voldoet aan :

$$\lambda_\mu = \frac{1}{\sqrt{N}} \mathbf{J} \cdot \boldsymbol{\eta}^\mu \geq \kappa_c \quad \forall \mu, \quad (4.19)$$

waarbij κ_c zo groot mogelijk moet zijn. Hierbij is κ_c genoteerd om aan te duiden dat het de minimale stabiliteit betreft van de kubische koppelingen, in tegenstelling tot κ_s en κ_b die reeds ingevoerd waren in de vorige hoofdstukken. Ik zal de oplossing van deze vergelijkingen het MSC perceptron¹ noemen. De precursor leerregel die de MSC tot resultaat heeft wordt dan gegeven door:

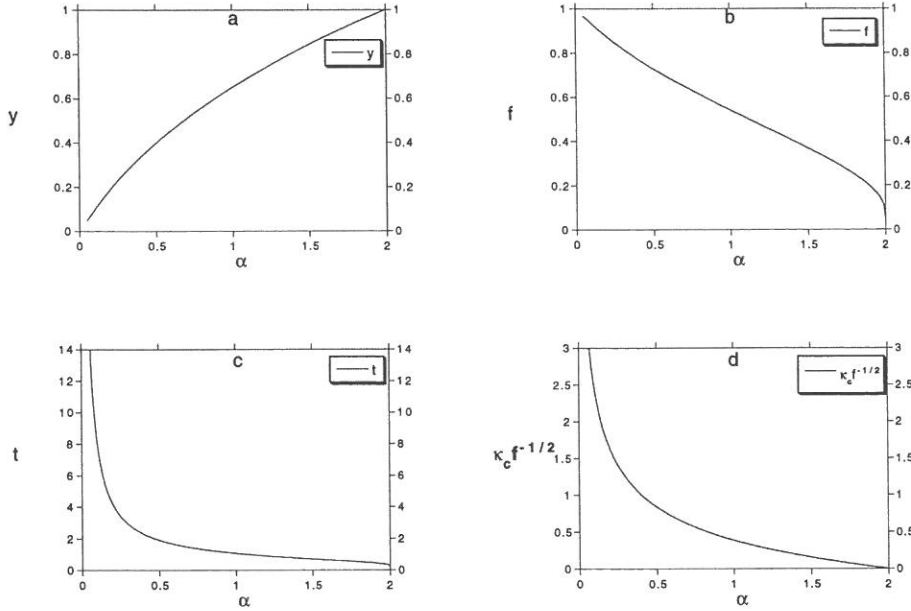
$$\begin{aligned} V(\lambda) &= (\kappa_c - \lambda)^2 \Theta(\kappa_c - \lambda) \\ P(J_i) &= \frac{1}{2} \Theta(1 - J_i^2), \end{aligned} \quad (4.20)$$

waarbij, zoals bij de MSN, de limiet voor $x \rightarrow \infty$ dient genomen te worden in de bekomen zadelpuntvergelijkingen om de maximaal stabiele oplossing te verzekeren [10].

De vier zadelpuntvergelijkingen ((4.3)...(4.6)) bepalen nu de drie overgebleven ordeparameters f , y en t alsook de stabiliteitsparameter κ_c , als functie van α . De numerieke oplossing van dit stelsel vergelijkingen is gegeven in figuur 4.2. Figuur 4.2a geeft de fractie van de componenten van de MSC vektor die in absolute waarde kleiner zijn dan 1. Deze fractie neemt toe van 0 bij $\alpha = 0$ — aandeuidend dat de MSC bij $\alpha = 0$ een binaire vektor is — tot 1 bij $\alpha = 2$. Dit laatste duidt aan dat de distributie (4.8) voor $P(j)$ bij $\alpha = 2$ een zuiver Gaussisch verloop heeft zonder de Dirac-delta pieken bij $j = \pm 1$, zoals ook het geval is voor de MSN. Figuur 4.2b toont de zelfoverlap f van de MSC vektor als functie van α . Er geldt dat $f = 1$ bij $\alpha = 0$, zoals verwacht wordt voor een binaire vektor. Deze zelfoverlap neemt verder af bij toenemende waarden van α als minder en minder componenten de binaire waarden ± 1 aannemen. Figuur 4.2c toont de "variantie" t van het Gaussische stuk van de distributie (4.8) voor $P(j)$. Bij kleine waarden van α is deze zeer groot zodat een groot gewicht

¹MSC staat voor "Maximally Stable Cubic".

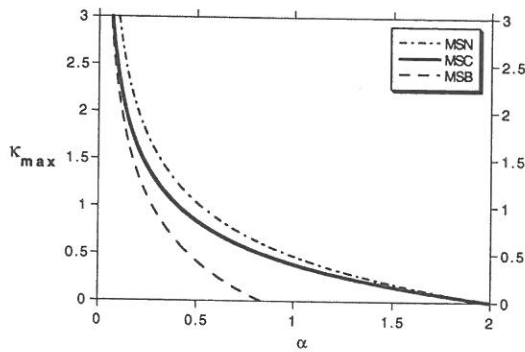
4. Leren in de hyperkubus



Figuur 4.2: Numerieke oplossing van de MSC vergelijkingen.

toegekend wordt aan de Dirac-delta pieken bij $j = \pm 1$. Bij toenemende waarden van α wordt t kleiner wat consistent met het kleiner aantal componenten dat binair is. Tenslotte toont figuur 4.2d de waarde van de minimale stabiliteit $\kappa_c/\sqrt{f}$ als functie van α . We hebben hier de "genormaliseerde" waarde van de stabiliteitsparameter weergegeven. Deze corrigeert voor de kortere lengte van de MSC vektor in vergelijking met de MSN en de MSB. Dit laat een zinvolle vergelijking toe tussen de MSC, de MSN en de MSB zoals weergegeven in figuur 4.3. Wanneer $\alpha \rightarrow 2$ gaat $\kappa_c/\sqrt{f} \rightarrow 0$ zoals ook geldig is voor de MSN. Dit is natuurlijk geen verrassend resultaat. Veel interessanter is echter het gedrag voor $\alpha \rightarrow 0$. In dit geval blijkt dat $\kappa_c/\sqrt{f}$ samenvalt met κ_b . De gehele indruk van figuur 4.3 is dat de MSC vloeiend interpoleert tussen de MSB (bij $\alpha = 0$) en de MSN (bij $\alpha = 2$). Deze indruk wordt nog versterkt als men terugkijkt naar figuur 4.2a voor $y(\alpha)$. Bij $\alpha = 0$ is $y = 1$ en dus is de MSC een binaire vektor, waar bij $\alpha = 2$ de distributie voor $P(j)$ terug Gaussisch is zoals bij

4.3. De MSC als Precursor



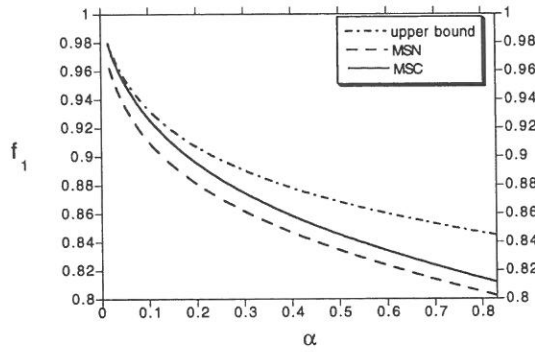
Figuur 4.3: De stabiliteitsparameter κ_{max} voor het maximaal stabiel perceptron bekomen met drie verschillende types van constraints: sferisch (MSN), kubisch (MSC) en binair (MSB). In het geval van de MSC is de herschaalde hyperkubus stabiliteit $\kappa_c/\sqrt{f}$ weergegeven om een zinvolle vergelijking te kunnen maken met de andere gevallen.

de MSN. Bij de saturatie limiet $\alpha = 0.83$ waar de MSB ophoudt te bestaan zijn er nog steeds 43% van de MSC componenten binair. Al deze algemene uitspraken versterken het vertrouwen in de MSC als een uitstekende precursor voor de MSB, in het bijzonder bij kleine waarden van α , maar geleidelijk aan verminderend in kwaliteit wanneer α toeneemt.

Figuur 4.4 toont de fractie van correct voorspelde componenten voor de MSB bij volledig knippen van de MSC precursor, de MSN en de bovengrens (4.18) zijn getoond om te vergelijken. Zoals verwacht levert de MSC een aanzienlijke verbetering voor kleine waarden van α . De bovengrens wordt zelfs dicht benaderd. Voor grotere waarden van α is de verbetering kleiner. Er is een wezenlijk verschil tussen de bovengrens en de MSC.

Figuur 4.5 toont de fractie $f_x(\alpha)$ van correct voorspelde synapsen van de MSB als enkel de xN grootste componenten van de MSC precursor geknipt worden. Curven zijn getoond voor het knippen van 40%, 60% en 80% ($x = 0.4$, 0.6 en 0.8) van de grootste componenten, en dit voor de MSC, de MSN en de bovengrens (4.17). Het bovenste setje van drie curven is voor $x = 0.4$, het

4. Leren in de hyperkubus



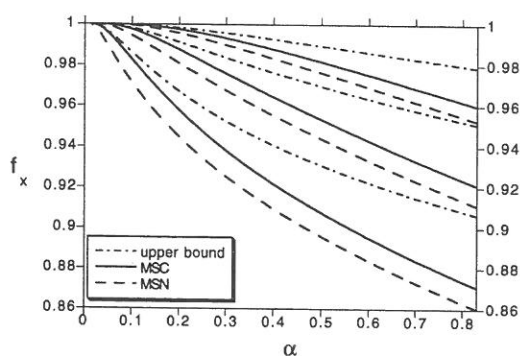
Figuur 4.4: De fractie $f_1(\alpha)$ van correct voorspelde synapsen van de MSB bij volledig knippen van de MSC. Het resultaat voor de MSN is gegeven als referentie alsook de bovengrens (4.18).

middenste voor $x = 0.6$ en het onderste setje is voor $x = 0.8$. Bemerk, zoals reeds vroeger opgemerkt, dat al deze curven gemaakt zijn met (4.15) voor een vaste waarde van J_o/t bepaald door $x = 2H(J_o/t)$. Analoog als voor figuur 4.4 levert de MSC een aanzienlijke verbetering voor kleine waarden van α waar hij de bovengrens zeer dicht benadert. Voor grotere waarden van α wordt de verbetering kleiner maar blijft er een wezenlijk verschil tussen de bovengrens en de MSC. Uit figuur 4.4 en 4.5 kunnen we concluderen dat de verandering van het minimalisatiedomein van de hypersfeer naar de ingeschreven hyperkubus voornamelijk een grote verbetering oplevert bij kleine waarden van α . In de volgende sectie zal ik tonen dat door een betere keuze van de potentiaal te maken ook een verbetering voor de grotere α waarden mogelijk wordt.

4.4 De optimale precursor in de hyperkubus

Ik zal in deze sectie de vraag beantwoorden of er betere keuzes bestaan dan de MSC als precursor potentiaal in de hyperkubus. Men zou kunnen trachten hierbij dezelfde weg in te slaan als in hoofdstuk 3. Helaas, een vergelijking afleiden voor de maximale overlap bereikbaar binnen de beschouwde klasse van kostfuncties op de hyperkubus bleek in dit geval niet mogelijk te zijn.

4.4. De optimale precursor



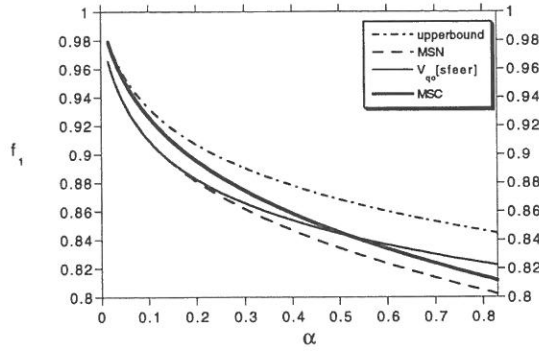
Figuur 4.5: De fractie $f_x(\alpha)$ van correct voorspelde synapsen van de MSB als enkel de xN grootste componenten van de MSC precursor geknipt worden. Hierbij geldt van boven naar onder respectievelijk $x = 0.4, 0.6$ en 0.8 . De resultaten voor de MSN alsook de bovengrens (4.17) zijn gegeven om te vergelijken.

Als alternatief wordt dan een potentiaal voorgesteld die de reeds gevonden bovengrens (4.17) zeer dicht benadert.

4.4.1 Een quasi-optimale potentiaal

Ik zal in deze sectie een potentiaal vooropstellen die de bovengrens $\hat{\gamma} \rightarrow 1$ zeer goed benadert. We hebben reeds een precursor gevonden die voor kleine waarden van α die bovengrens uitstekend benadert: de MSC. Vermits de gebruikte potentiaal bij de MSC dezelfde is als bij de MSN was de verbetering van de MSC t. o. v de MSN volledig te wijten aan de keuze van de hyperkubus als minimalisatiedomein. We kunnen dus nu trachten door ook de potentiaal te wijzigen een precursor te vinden die ook voor grotere waarden van α de bovengrens zeer goed benadert. Een suggestie volgt uit figuur 4.6. De quasi-optimale potentiaal op de hypersfeer levert voor kleine waarde van α een bijna identiek resultaat als de MSN, maar voor grotere waarden van α wordt hij geleidelijk aan beter tot hij zelfs beter wordt dan de MSC. Deze potentiaal lijkt dus een goede kandidaat om te gebruiken in de hyperkubus, vermits hij

4. Leren in de hyperkubus



Figuur 4.6: De fractie van correct geknipte componenten bij volledig knippen van de precursor. Resultaten zijn gegeven voor de MSN, de MSC en de quasi-optimale potentiaal $V_{qo}(\lambda)$ op de hypersfeer. De bovengrens (4.18) is weergegeven met een punt-streep lijn.

net daar superieur is waar de MSC het zwakst is: bij de grotere waarden van α . Dus laten we de resultaten eens bekijken in de hyperkubus met gebruik van de quasi-optimale potentiaal² uit (3.17):

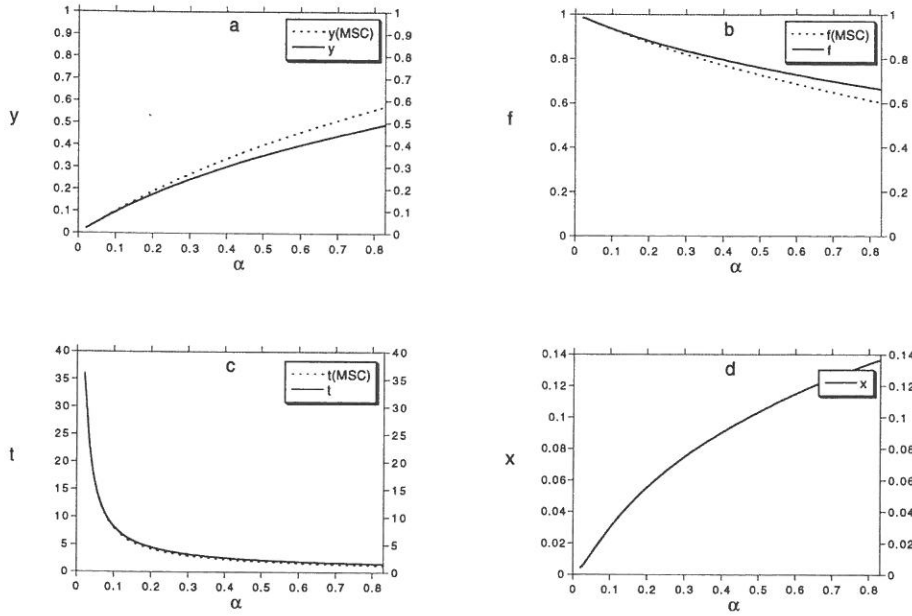
$$V_{qo}(\lambda) \equiv \begin{cases} \frac{1}{\lambda - \kappa_b} & : \text{als } \lambda \geq \kappa_b \\ +\infty & : \text{als } \lambda < \kappa_b. \end{cases} \quad (4.21)$$

Figuur 4.7 geeft de numerieke oplossing weer van de zadelpuntvergelijkingen ((4.3)...(4.6)) met $V(\lambda)$ gegeven door (4.21) voor $\alpha \in [0, 0.83]$, wat het relevante oplossingsgebied is voor een MSB precursor. Bij figuren (4.7a...4.7c) is tevens de oplossing van de MSC toegevoegd om te vergelijken. De algemene tendens van de oplossing is dezelfde als in het geval van de MSC. Er blijkt duidelijk uit figuur 4.7a en 4.7b dat de quasi-optimale potentiaal voor de grotere α -waarden een sterkere graad van binairheid heeft dan de MSC.

Figuur 4.8 toont de fractie (4.16) van correct voorspelde componenten van de MSB bij volledig knippen van de precursor bekomen met de quasi-optimale

²Noteer dat deze potentiaal een kostfunctie definieert die convex is in de hyperkubus. Een bewijs hiervan is gegeven in bijlage D. De uniciteit van het minimum is dus verzekert.

4.4. De optimale precursor

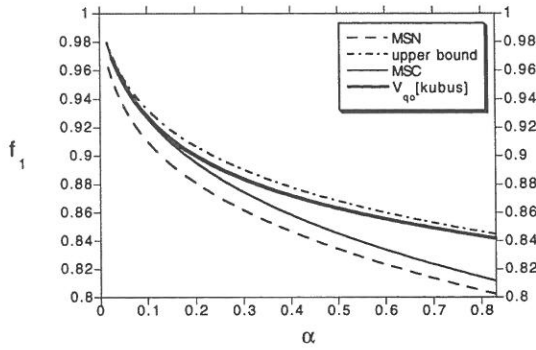


Figuur 4.7: Numerieke oplossing van de vergelijkingen voor de quasi-optimale potentiaal in de hyperkubus voor $\alpha \in [0, 0.83]$. De oplossing voor de MSC is bij $y(\alpha)$, $f(\alpha)$ en $t(\alpha)$ toegevoegd om te vergelijken.

potentiaal (4.21). De reeds eerder vermelde resultaten voor de MSN en de MSC, alsook de bovengrens (4.18) zijn gegeven om te vergelijken. Zoals verwacht is er een opmerkelijke verbetering voor de grote α waarden. Bij $\alpha = 0.83$ stijgt de fractie van correct geknipte componenten met ongeveer 3% van 81% voor de MSC tot 84% voor de quasi-optimale potentiaal in de hyperkubus. Deze potentiaal benadert over het hele gebied van α waarden de bovengrens (4.18) zeer goed. Het is de best gevonden precursor binnen de beschouwde klasse van kostfuncties.

Numerieke simulaties bevestigen de superioriteit van de precursor bepaald door de quasi-optimale potentiaal in de hyperkubus. Figuur 4.9 toont de minimale stabiliteit $\kappa(\alpha)$ van de binaire vektor bekomen door volledig knippen van

4. Leren in de hyperkubus

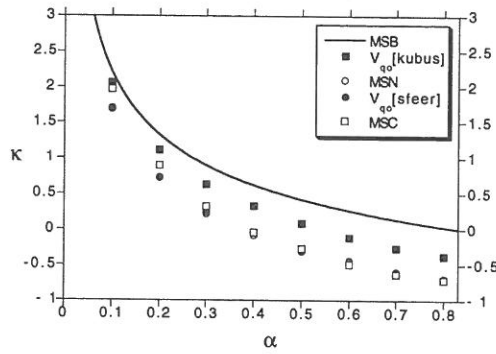


Figuur 4.8: Fractie van binaire synapsen in de MSB die correct voorspeld worden door het knippen van alle componenten van de quasi-optimale potentiaal V_{qo} . De reeds eerder getoonde resultaten voor de MSN en de MSC zijn toegevoegd om te vergelijken. De punt-streep lijn geeft de bovengrens (4.18) weer.

deze precursor voor een perceptron met 50 input neuronen. Als vergelijking is tevens de minimale stabiliteit getoond die bekomen is door volledig knippen van de componenten van de MSC, de MSN en de quasi-optimale hypersfeer precursor bepaald door V_{qo} uit (3.17).

De quasi-optimale hyperkubus precursor komt zeer goed uit de simulaties in vergelijking met de anderen en dit over het gehele gebied van α waarden. Het vermindert de afstand tussen de theoretische waarden $\kappa_b(\alpha)$ en de MSN met meer dan de helft. Het resultaat voor de andere twee precursors ligt er tussenin. Zoals verwacht liggen de resultaten voor de MSC, voor kleine waarden van α , dicht bij die van de quasi-optimale hyperkubus precursor. De MSC verslechtert echter snel voor grotere waarden van α waar het uiteindelijk samenvalt met de MSN. Het resultaat voor de quasi-optimale potentiaal op de hypersfeer valt samen met de MSN voor kleine α waarden en is slechts een beetje beter bij grotere waarden van α . Deze simulaties bevestigen de hogere kwaliteit van de quasi-optimale potentiaal op de hyperkubus en benadrukken dat zowel de hyperkubus constraint als de quasi-optimale potentiaal essentieel zijn voor dit uitstekend resultaat.

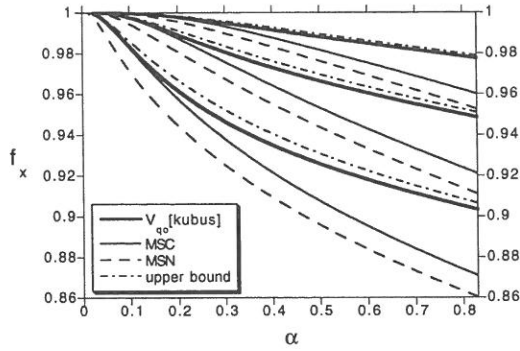
4.4. De optimale precursor



Figuur 4.9: De minimale stabiliteit κ voor een perceptron met $N = 50$ uit numerieke simulaties. De binaire synapsen worden bekomen door volledig knippen van vier verschillende precursors. Bij kleine waarden van α vallen de resultaten voor de MSN samen met die van de potentiaal V_{qo} op de hypersfeer, en bij grote waarden van α met die van de MSC.

Ondanks deze grote verbetering voorspelt zelfs de beste hyperkubus precursor — bepaald door de bovengrens (4.18) — 16% van de MSB gewichten verkeerd nabij $\alpha = 0.83$. Een bijkomend leerproces waarin deze componenten geïdentificeerd en verbeterd worden is dan ook noodzakelijk. Opnieuw verwachten we dat de sterkste componenten de meest betrouwbare voorspellers zijn en de zwakste de meest twijfelachtige. Dit kunnen we natuurlijk weer bevestigen door te kijken naar de kwaliteitsmaat $f_x(\alpha)$. In figuur 4.10 is de fractie f_x uit (4.15) gegeven voor drie waarden van J_o/t corresponderend met $x = 0.4$, $x = 0.6$ en $x = 0.8$. Voor elke van deze gevallen is tevens de bovengrens (4.17) weergegeven. De figuur toont de merkbaar hogere kwaliteit van de quasi-optimale potentiaal in de hyperkubus. Vergeleken met het resultaat voor de MSN is er een wezenlijke verbetering voor alle waarden van α . Voor kleine waarden van α is dit weer volledig te wijten aan de hyperkubus constraint vermits de verbetering ook bekomen wordt voor de MSC. Voor grotere waarden van α komt de hogere kwaliteit van de quasi-optimale hyperkubus precursor sterk naar voren. De combinatie van de hyperkubus constraint met

4. Leren in de hyperkubus



Figuur 4.10: De fractie van binaire componenten van de MSB die correct voorspeld zijn door het knippen van 40% ($x = 0.4$) (boven), 60% ($x = 0.6$) (midden) en 80% ($x = 0.8$) (onder) van de synapsen van de quasi-optimale hyperkubus precursor bepaald door V_{qo} . De reeds eerder getoonde resultaten voor de MSN en de MSC zijn toegevoegd om te vergelijken. De punt-streep lijn toont de bovengrens (4.17)

de quasi-optimale potentiaal V_{qo} levert daar sterk verbeterde resultaten. De numerieke waarde van $f_x(\alpha)$ duidt erop dat de 40% sterkste componenten van de quasi-optimale hyperkubus precursor zeer betrouwbare voorspellers zijn van de overeenkomstige MSB synapsen. De waarschijnlijkheid om een foute voorspelling te maken is klein wanneer 60% van de sterkste synapsen geknipt worden. Ze bedraagt ongeveer 5% bij $\alpha = 0.83$. Bij het knippen van de 80% sterkste synapsen zijn er ongeveer 9% foute voorspellingen bij dezelfde α -waarde. Dit betekent dat er van de $0.8N$ componenten er $(0.09)(0.8N) = 0.072N$ foute componenten zijn. Als we vergelijken met figuur 4.8 waar er een totaal van ongeveer $0.16N$ foute componenten voorspeld wordt, moeten we besluiten dat de overige $(0.16 - 0.072)N = 0.088N$ foute voorspellers bij de 20% zwakste componenten zitten. De grootste concentratie van fouten zit dus bij de 20% zwakste synapsen. Een tweede leerproces moet zich dus voornamelijk op deze componenten concentreren om de bekomen binaire vektor alsnog te verbeteren.

4.5 Simulaties

Net zoals in vorig hoofdstuk zullen we hier trachten met simulaties te onderzoeken wat de maximaal mogelijke stabiliteit is die nog bekomen kan worden na het knippen van een fractie van de grootste componenten van de quasi-optimale hyperkubus precursor. Dit geeft dus een indicatie van wat de invloed is van de verkeerd geknipte componenten op de κ waarde. Om dit te onderzoeken is het weer noodzakelijk van de nog niet geknipte componenten te bepalen met een enumeratie. Zoals in vorig hoofdstuk is de gebruikte enumeratie methode het Gray-code algoritme [30].

De simulaties verlopen op vrijwel identieke wijze als in hoofdstuk 3.

1. Kies de grootte van het netwerk, N .
2. Kies de waarde van $\alpha = p/N$.
3. Kies een set van $p = \alpha N$ patronen $\boldsymbol{\eta}^\mu$. De simulaties zijn uitgevoerd met Gaussische patronen om finite size effecten te verkleinen [22].
4. Bepaal vervolgens de gewenste precursor. Voor het bepalen van de MSN werd het AdaTron algoritme [2] gebruikt. In het geval van de quasi-optimale precursor op de hyperkubus werd gebruik gemaakt van een standaard gradient-descent algoritme. Deze minimalisatie van de kostfunctie is snel en eenvoudig omdat de kostfunctie een uniek minimum heeft op de hyperkubus.
5. Zodra de precursor berekend is kunnen we ze zijn componenten in absolute waarde rangschikken van groot naar klein. Dit zal straks het knip proces vereenvoudigen.
6. Het is noodzakelijk van dezelfde herrangschikking van de componenten uit te voeren op de patronen $\boldsymbol{\eta}^\mu$ zodat de scalaire produkten $\mathbf{J} \cdot \boldsymbol{\eta}^\mu$ en dus de stabiliteiten dezelfde waarde blijven behouden.
7. Knip een zeker percentage van de grootste componenten. Vermits met het gebruikte Gray-code algoritme enumereren van meer dan 25 componenten tijdrovend wordt, mag het aantal componenten dat overblijft na knippen de 25 niet overschrijden. Deze geknipte componenten worden tijdens de overige stappen in het algoritme vast gehouden, hierbij veronderstellend dat deze componenten betrouwbare voorspellers zijn van de

4. Leren in de hyperkubus

overeenkomstige MSB synapsen. Het hoofddoel van de simulaties is van dit laatste te verifiëren.

8. De overige niet geknipte componenten worden vrij gehouden om deel te nemen aan een volledige enumeratie met het "Gray-code" algoritme. Tijdens de enumeratie wordt de binaire vektor met de grootste minimale stabiliteit κ_b geselecteerd als benadering voor de MSB.
9. Ga terug naar stap 3 en herhaal de volgende stappen in het algoritme. Doe dit 200 maal.
10. Bereken het gemiddelde van de 200 bekomen κ_b waarden.

Dit algoritme dient herhaald te worden voor verschillende waarden van α . De gevonden gemiddelde waarden van κ_b worden dan uitgezet als functie van de parameter α .

In de quasi-optimale hyperkubus precursor zijn er vele componenten ± 1 , zeker voor kleine waarden van α . (Zie figuur 4.7a) Dit levert een probleem als er meer componenten van de precursor ± 1 zijn dan we wensen te knippen. Hoe moet men dan de sterkste componenten identificeren? Voor dat ik de resultaten van de simulaties zal bespreken zal ik in volgende sectie dit probleem aanpakken.

4.5.1 Oplossing voor de ± 1 ontaarding.

We kunnen aan de kostfunctie $E(\mathbf{J})$ een "perturbatie" toevoegen die de ± 1 ontaarding gedeeltelijk oplost. Een voor de hand liggende keuze voor de perturbatie is de term $\rho \mathbf{J}^2$ met $\rho > 0$. Deze term oefent duidelijk een kracht uit die het minimum van de kostfunctie naar de oorsprong $\mathbf{J} = 0$ toetrekt. Nauwkeurige regeling van de parameter ρ maakt het mogelijk van de fractie $1 - y$ van ± 1 componenten te doen afnemen tot op een gekozen waarde. De waarde die ρ hiertoe moet aannemen kan eenvoudig berekend worden door de term $\rho \mathbf{J}^2$ toe te voegen aan de kostfunctie $E(\mathbf{J})$. De zadelpuntvergelijkingen voor de aldus bekomen precursor in de hyperkubus worden dan :

$$y - 2\rho x = -\alpha \int_{-\infty}^{+\infty} D\sigma \left[\frac{\partial \lambda_o}{\partial \sqrt{f}\sigma}(\sqrt{f}\sigma, x) - 1 \right], \quad (4.22)$$

$$(ty)^2 = \alpha \int_{-\infty}^{+\infty} D\sigma \left[\lambda_o(\sqrt{f}\sigma, x) - \sqrt{f}\sigma \right]^2, \quad (4.23)$$

$$y = \text{Erf}\left(\frac{1}{t\sqrt{2}}\right), \quad (4.24)$$

$$f = 1 - y + y t^2 - \sqrt{\frac{2}{\pi}} t \exp\left[\frac{-1}{2t^2}\right]. \quad (4.25)$$

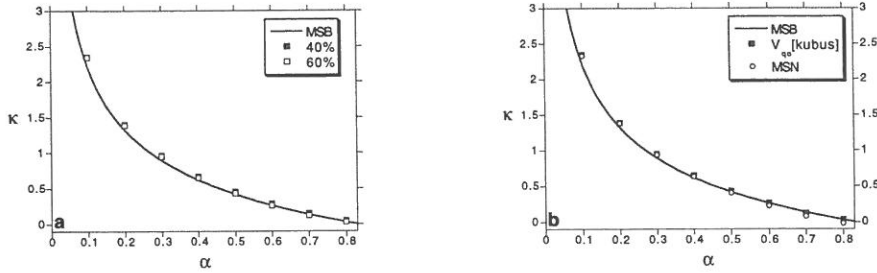
Een gedetailleerde beschrijving van de afleiding van deze vergelijkingen is gegeven in bijlage C in sectie C.5. Dus alleen de zadelpuntvergelijking (4.3) is gewijzigd. Voor een gegeven waarde van α zijn we nu vrij de fractie $1 - y$ van ± 1 componenten te kiezen. Deze moet natuurlijk wel een kleinere waarde aannemen dan de waarde voor $\rho = 0$. Bij een keuze van y bepalen de vier zadelpuntvergelijkingen dan de parameters f , t , x en ρ . Noteer dat de extra term ρJ^2 de convexiteit van de kostfunctie niet beïnvloed wanneer $\rho > 0$, zodat een uniek minimum van de kostfunctie in de hyperkubus verzekert is. (Zie ook bijlage D.)

Indien er dus meer componenten ± 1 zijn dan men wenst te knippen in stap 7 van het algoritme doet men een extra minimalisatie. Men minimaliseert de kostfunctie opnieuw met dezelfde set van patronen maar ditmaal met de ρJ^2 term en met een geschikte keuze voor ρ . Deze laatste moet zodanig gekozen zijn dat het aantal ± 1 componenten gereduceerd is tot een getal dat gelijk is aan of kleiner is dan het aantal te knippen componenten. Het is nu mogelijk van de sterkste te knippen componenten te identificeren in de hyperkubus precursor bekomen met $\rho = 0$.

4.5.2 Resultaten

Figuren 4.11 en 4.12 tonen de resultaten van de numerieke simulaties. De minimale stabiliteit κ is weergegeven als een functie van α voor de beste binaire vektor bekomen met de quasi-optimale hyperkubus precursor. Elk data punt vertegenwoordigt een gemiddelde over 200 steekproeven. De input patronen zijn random Gaussische patronen [22]. Figuur 4.11 toont de resultaten voor een relatief klein systeem $N = 40$. Figuur 4.11a toont het effect van het knippen van verschillende fracties van de quasi-optimale hyperkubus precursor. Wanneer enkel 16 (40%) van de precursor componenten geknipt worden voorspellen de theoretische curven uit figuur 4.10 dat alle geknipte componenten met een grote waarschijnlijkheid de correcte MSB synaps opleveren. De simulaties bevestigen deze voorspelling heel goed. De numerieke waarde van κ ligt zelfs hoger dan de theoretische curve $\kappa_b^{max}(\alpha)$. Wanneer 24 (60%)

4. Leren in de hyperkubus

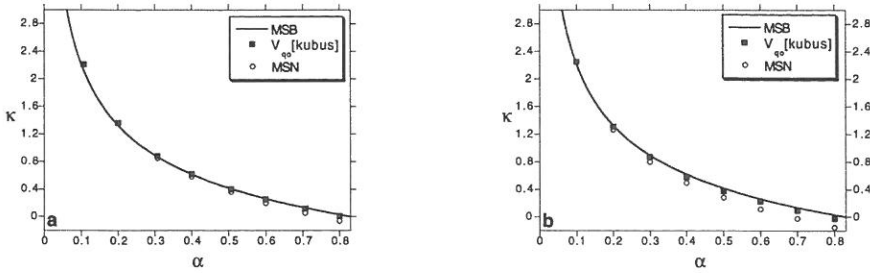


Figuur 4.11: De minimale stabiliteit κ voor een perceptron met $N = 40$. Figuur (a) toont de waarde van κ wanneer 40% of 60% van de componenten van de quasi-optimale hyperkubus precursor geknipt worden. Figuur (b) vergelijkt de MSN en de quasi-optimale hyperkubus precursor wanneer de sterkste 60% componenten geknipt worden. De volle lijn toont zoals gewoonlijk de theoretische curve $\kappa_b^{max}(\alpha)$.

van de componenten geknipt worden, leiden we af uit figuur 4.10 dat, voor grote waarden van α , tenminste één van de geknipte componenten een incorrecte MSB synaps zou kunnen voorspellen. De numerieke simulaties blijven goede resultaten opleveren voor alle waarden van α maar de resultaten voor 60% knippen liggen een beetje lager dan die van 40% knippen. Figuur 4.11b vergelijkt de resultaten van twee verschillende precursors: de quasi-optimale hyperkubus precursor en de MSN precursor³. In beide gevallen zijn de 60% sterkste componenten geknipt. Ook al wordt er voor de MSN een verdubbeling van incorrecte binaire synapsen verwacht, de resultaten van de numerieke simulaties blijven, voor kleine waarden van α , mooi overeenstemmen met de theoretische curve $\kappa_b^{max}(\alpha)$ maar de fit verslechtert een beetje voor de grotere α waarden.

Figuur 4.12 toont resultaten voor grotere netwerken $N = 75$ en $N = 100$. In deze gevallen moet er een veel groter aantal componenten geknipt worden omdat het tijdsgebruik van het Gray-code algoritme ons beperkt tot een enumeratie van 25 synapsen. Dit heeft tot gevolg dat veel meer geknipte componenten

³In al de volgende simulaties zullen geen resultaten getoond worden voor de quasi-optimale potentiaal op de hypersfeer vermits het verschil met de MSN klein is.



Figuur 4.12: De minimale stabiliteit κ voor een perceptron met $N = 75$ (a) en $N = 100$ (b). Beide figuren vergelijken de MSN en de quasi-optimale hyperkubus precursor. Er zijn enkel 25 componenten geënumereerd zodat er een aanzienlijk aantal fout geknipte synapsen verwacht wordt. De overeenstemming met de theoretische curve blijft desalniettemin goed. De volle lijn toont zoals gewoonlijk de theoretische curve $\kappa_b^{max}(\alpha)$.

een foute voorspelling opleveren voor hun overeenkomstige MSB synaps. Voor de quasi-optimale potentiaal kan men uit figuur 4.10 schatten hoeveel foute voorspellingen er zijn. In het geval $N = 75$ blijken er ongeveer 3 synapsen fout geknipt te worden bij grote waarden van α en voor $N = 100$ zijn er dat 6. Het interessante punt is nu te onderzoeken hoe deze foute componenten de waarde van de minimale stabiliteit beïnvloeden. Voor $N = 75$ met 50 sterkste componenten van de quasi-optimale hyperkubus precursor geknipt blijken de simulatieresultaten verrassend genoeg heel mooi de theoretische curve $\kappa_b^{max}(\alpha)$ te volgen, en dit over het hele gebied van α waarden. Voor $N = 100$ met 75 synapsen geknipt blijft de overeenstemming met $\kappa_b^{max}(\alpha)$ uitstekend voor kleine waarden van α en de afwijking bij grote waarden van α is klein. Dit onverwacht resultaat geeft aan dat de verschillende incorrecte synapsen bekomen door het knippen van de quasi-optimale hyperkubus precursor de hoge waarde van de stabiliteit niet vernietigen maar enkel een kleine vermindering van zijn waarde veroorzaken. Blijkbaar is de waarde van deze componenten niet cruciaal om een grote waarde te bekomen van de minimale stabiliteit. Dit is een zeer bevredigend resultaat omdat het toont dat de hyperkubus precursor zelfs betere resultaten oplevert dan verwacht kan worden uit figuur 4.10. De resul-

4. Leren in de hyperkubus

taten bekomen voor de MSN zijn getoond om te kunnen vergelijken. Volgens figuur 4.10 zijn er in dit geval nog een groter aantal incorrecte synapsen. De MSN levert echter ook mooie resultaten voor kleine waarden van α maar doet het opmerkelijk minder goed voor grotere α waarden.

Hoofdstuk 5

Optimale leerstrategie voor het binair perceptron

5.1 Inleiding

In dit laatste hoofdstuk zal ik een strategie introduceren die de exacte MSB als oplossing heeft voor perceptrons tot $N = 50$. We wensen dus de binaire vektor $\mathbf{B}$ te vinden die voldoet aan:

$$\Lambda_\mu \equiv \frac{1}{\sqrt{N}} \mathbf{B} \cdot \boldsymbol{\eta}^\mu \geq \kappa_b > 0 \quad \forall \mu (\mu = 1 \dots p), \quad (5.1)$$

met κ_b zo groot mogelijk. Zoals reeds vermeld in hoofdstuk 1 is het vinden van een MSB vektor een NP-compleet probleem. Het enige algoritme wat gekend is om het probleem op te lossen is een enumeratie van alle mogelijke combinaties van ± 1 koppelingen. De oplossingsstrategie is dus van alle mogelijke 2^N binaire vectoren $\mathbf{B}$ te exploreren en die vektor te selecteren die de grootste waarde heeft voor de minimale stabiliteit κ_b . De eerste vraag die we moeten beantwoorden is hoe we al die mogelijke vectoren gaan opsommen. Een eerste manier om dit te doen is het Gray-code algoritme. Een andere methode is het branch en bound algoritme. Dit is echter geen enumeratie techniek in de letterlijke zin van het woord. Niet alle vectoren worden in deze methode doorlopen, maar het algoritme is zodanig dat diegene die niet doorlopen worden ook onmogelijk de oplossing kunnen zijn van het probleem. Men bespaart dus kostbare computertijd door bepaalde vectoren niet te beschouwen waarvan men zeker is dat ze de oplossing niet kunnen zijn. Het gevolg is dat het resultaat van een branch en bound techniek hetzelfde is als dat van een volledige enumeratie,

5. Optimale leerstrategie voor het binair perceptron

maar dat ze een stuk sneller verloopt. Omdat enumereren een computertijd vergt die exponentieel toeneemt met de systeemgrootte N , zijn dergelijke methodes in de praktijk enkel haalbaar voor voldoende kleine systemen. Indien men het Gray-code algoritme gebruikt om de enumeratie uit te voeren kan men perceptrons aan met $N \leq 30$ [5, 22, 24]. Recent hebben Milde en Kobe [23] een branch en bound algoritme [30] gebruikt waarmee ze *exacte* oplossingen bekomen hebben voor het MSB perceptron voor systemen tot $N = 40$. Dit is een imponerende verbetering als men zich realiseert dat het aantal mogelijke binaire perceptrons die moeten geëxploreerd worden, toeneemt met een factor 2^{10} in vergelijking met de vorige systeemgrootte van $N = 30$.

Het idee is nu van de "knip-techniek" uit vorige hoofdstukken te combineren met de branch en bound methode die de vroegere enumeratietechniek ("Gray-code") vervangt, om zo oplossingen te bekomen voor het MSB perceptron tot $N = 50$. Vooraleer de gebruikte strategie in meer detail te behandelen zal ik eerst de werking van het branch en bound algoritme uiteenzetten.

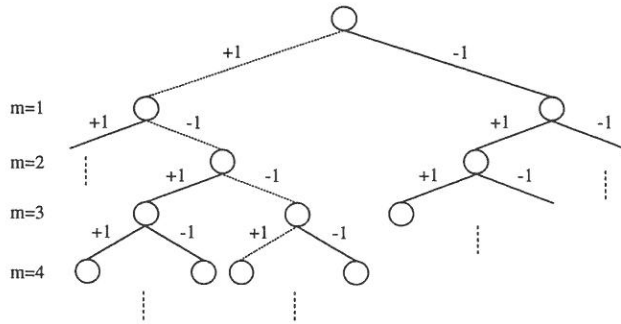
5.2 De methode van branch en bound

De manier waarop branch en bound alle mogelijke binaire vectoren $\mathbf{B}$, voor een perceptron met N koppelingen, exploreert is gebaseerd op een binaire boom met N niveau's ($m = 1, m = 2, \dots, m = N$). (Zie figuur 5.1)

Elke knoop van de boom vertakt in twee andere knopen. Deze twee vertakkingen representeren de twee mogelijke binaire waarden van een component van $\mathbf{B}$. Indien men de vertakkingen van de boom volgt van boven (vanuit het startpunt) naar beneden (stijgende m -waarden) moet men bij elke knoop een van de twee vertakkingen kiezen. Hiermee wordt dan de waarde van de volgende component van $\mathbf{B}$ vastgelegd. (Indien men zich op niveau $m = 1$ bevindt is B_1 bepaald, op niveau $m = 2$ is B_1 en B_2 bepaald, enz. ...) Op het laagste niveau van de boom, bij $m = N$ heeft men dan een mogelijke binaire vektor geconstrueerd. Er zijn zo 2^N mogelijke "wegen" van boven naar beneden die men kan volgen. Dit is het "branching" deel van het algoritme.

Terwijl men de boom afdaalt via een gekozen weg wordt er een getal berekend bij elke knoop. Dit getal is een bovengrens voor de nog maximaal mogelijke stabiliteit die nog bereikbaar is na het reeds gevolgde deel van de weg. Veronderstel even dat we helemaal bovenaan zitten in de boom, op $m = 0$. Er is dan nog geen enkele component gekozen omdat er nog geen vertakkingen gevolgd zijn. Een mogelijke bovengrens voor de nog maximaal mogelijke stabiliteit is

5.2. De methode van branch en bound



Figuur 5.1: Een voorstelling van een deel uit een binaire boom. De vertakkingen in puntlijn stellen een mogelijke "weg" voor. De keuzes die gemaakt worden langsheen deze "weg" resulteren in een binaire vektor waarvan de eerste 4 componenten gegeven worden door $(+1, -1, -1, +1, \dots)$

dan $\text{Min}_\mu \left[\frac{1}{\sqrt{N}} \sum_j |\eta_j^\mu| \right]$. (Dit is duidelijk een bovengrens omdat het de minimale stabiliteit is als $\mathbf{B} = \text{sign}(\boldsymbol{\eta}^\mu)$ wat natuurlijk alleen maar kan als $p = 1$.) Als we verder afdalen in de boom dan kunnen we dus, bij een knoop op niveau m , als mogelijke bovengrens nemen:

$$\text{Min}_\mu \left[\frac{1}{\sqrt{N}} \left(\sum_j^N |\eta_j^\mu| - 2 \sum_j^m |\eta_j^\mu| \Theta(-B_j \eta_j^\mu) \right) \right], \quad (5.2)$$

vermits de eerste m componenten B_j van de vektor gekozen zijn. Als men de boom afdalt van knoop tot knoop zal deze bovengrens telkens opnieuw aangepast worden aan de nieuwe gemaakte keuzen. Merk op dat, als $m = N$, (5.2) gelijk is aan de minimale stabiliteit voor de bekomen binaire vektor op de bodem van de boom. De reden waarom het branch en bound algoritme in staat is grotere systemen te enumereren dan het Gray-code algoritme, is omdat het eerste algoritme opzettelijk een aantal vertakkingen niet exploreert. Dit is het "bounding" deel van het algoritme. Natuurlijk moet men —indien men een exacte oplossing wenst— zeker zijn dat de oplossing zich niet bevindt in de niet geëxploreerde vertakkingen van de boom. Dit wordt verzekert door vooraf een benaderde oplossing te construeren van (5.1). Ik zal deze oplossing een κ_b -schatting noemen. Een mogelijke κ_b -schatting is de minimale stabiliteit bekomen door een precursor volledig te knippen. Bij elke

5. Optimale leerstrategie voor het binair perceptron

knoop wordt de hierboven berekende bovengrens (5.2) nu vergeleken met de κ_b -schatting. Als de bovengrens kleiner is dan die schatting wordt de vertakking niet verder geëxploreerd omdat de oplossing die bekomen zou worden door die vertakking wél te volgen een kleinere minimale stabiliteit zal hebben dan de reeds bekomen κ_b -schatting. Na zulk een wegsnijden van een vertakking wordt teruggekeerd naar de vorige knoop die nog een ongeëxploreerde vertakking heeft. Als er gedurende het algoritme een "weg" is via dewelke men de bodem van de boom bereikt voordat alle vertakkingen onderzocht zijn, wordt de κ_b -schatting vervangen door de minimale stabiliteit van de binaire vektor bekomen via die "weg". Met deze nieuwe schatting worden dan de andere alsnog niet geëxploreerde vertakkingen onderzocht. Op het einde van het algoritme zal de binaire vektor met de grootste waarde van de minimale stabiliteit de MSB zijn.

5.2.1 Hoe efficiënt programmeren?

Berekening van de bovengrens

Uit voorgaande beschrijving van het branch en bound algoritme blijkt dat we bij elke knoop, volgens formule (5.2), p getallen dienen te berekenen waarvan het kleinste dient geselecteerd te worden als bovengrens. Het berekenen van deze p getallen is het meest tijdconsumerende deel van het algoritme. Indien bij elke knoop op niveau m de volledige sommatie gegeven door $\sum_j^N |\eta_j^\mu| - 2 \sum_j^m |\eta_j^\mu| \Theta(-B_j \eta_j^\mu)$, p keer moet berekend worden zal het algoritme ontzettend lang duren. We moeten trachten van de berekening van (5.2) te versnellen. Dit kan op eenvoudige wijze gedaan worden. Noteer om te beginnen de p te berekenen getallen in een knoop op niveau m als S_m^μ , dus:

$$S_m^\mu = \sum_j^N |\eta_j^\mu| - 2 \sum_j^m |\eta_j^\mu| \Theta(-B_j \eta_j^\mu). \quad (5.3)$$

Bekijk nu even de grootheid $\sum_j^m B_j \eta_j^\mu$, dit kan geschreven worden als:

$$\sum_j^m B_j \eta_j^\mu = \sum_j^m |\eta_j^\mu| - 2 \sum_j^m |\eta_j^\mu| \Theta(-B_j \eta_j^\mu). \quad (5.4)$$

Gebruik makend van deze gelijkheid kan men (5.3) herschrijven als:

$$S_m^\mu = \sum_j^N |\eta_j^\mu| - \sum_j^m |\eta_j^\mu| + \sum_j^m B_j \eta_j^\mu. \quad (5.5)$$

5.2. De methode van branch en bound

Deze uitdrukking kan nog herschreven worden als men de waarde van S^μ kent op niveau $m - 1$:

$$S_m^\mu = S_{m-1}^\mu - |\eta_m^\mu| + B_m \eta_m^\mu. \quad (5.6)$$

Hieruit volgt onmiddellijk dat men zeer vlug de p getallen S_m^μ op niveau m in de boom kan berekenen als men de p getallen S_{m-1}^μ kent van de vorige knoop $m - 1$. Op deze manier moet men dan de sommaties die optreden in S_m^μ niet telkens opnieuw berekenen. Natuurlijk kon men uitdrukking (5.3) direkt schrijven m. b. v. de informatie van de p getallen uit de vorige knoop op niveau $m - 1$. De reden waarom we (5.3) omgevormd hebben tot (5.5) is omdat dan de Θ -functie niet meer voorkomt. Om de Θ -functie te programmeren moet men gebruik maken van IF-THEN constructies en deze vragen meer computertijd dan een optelling of een vermenigvuldiging.

Struktuur van het programma

In deze korte sectie zal ik zeer bondig de structuur schetsen van de subroutine die het branch en bound algoritme uitvoert. Het is een recursieve subroutine. Dit wil zeggen dat de subroutine zichzelf oproept. Vermits een subroutine pas eindigt als alle programma statements doorlopen zijn, zal de subroutine pas stoppen als alle recursief opgeroepen branch en bound subroutines afgehandeld zijn. Op een willekeurig moment tijdens de uitvoering van het branch en bound algoritme zullen er dus een aantal geneste branch en bound routines tegelijk lopen. Deze techniek maakt het mogelijk op eenvoudige wijze het branch en bound algoritme te programmeren. Hieronder is op schematische wijze de recursieve subroutine beschreven.

Recursieve subroutine SEARCH(B, m, S_{m-1}^μ)¹

- Bepaal de p getallen S_m^μ met behulp van S_{m-1}^μ .
- Bepaal $\text{Min}_\mu S_m^\mu$, de gezochte bovengrens.
- IF (bovengrens < κ_b -schatting) THEN
 RETURN²
 ELSE

¹ S_{m-1}^μ zijn de p getallen uit de vorige knoop $m - 1$.

²De subroutine wordt verlaten en er wordt teruggekeerd naar de subroutine van de vorige knoop.

5. Optimale leerstrategie voor het binair perceptron

```
CALL SEARCH(B,m+1, $S_m^\mu$ )  
  B(m+1) = - B(m+1)3  
,  
  CALL SEARCH(B,m+1, $S_m^\mu$ )  
ENDIF
```

- Als men de bodem van de boom bereikt en dus een eerste binaire vektor gevonden heeft:
IF ($\kappa^4 > \kappa_b$ -schatting) THEN
 κ_b -schatting = κ^5
 Bewaar tevens de gevonden binaire vektor
ELSE
 RETURN⁶
ENDIF

Als de branch en bound routine stopt is de vektor met de grootste minimale stabiliteit gekend.

5.3 De leerstrategie

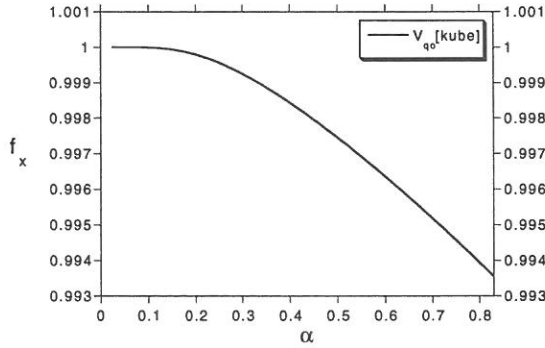
De leerstrategie kan in twee grote delen gesplitst worden. Het eerste deel bestaat uit het berekenen van de quasi-optimale hyperkubus precursor. Van deze precursor wordt dan een deel van de componenten geknipt, waarvan geweten is dat ze met zeer grote waarschijnlijkheid de overeenkomstige synaps van de MSB correct voorspellen. Het tweede deel bestaat er dan in van de overige niet geknipte componenten te bepalen met de methode van branch en bound. Door een gedeelte van de componenten van de precursor te knippen en de andere te onderwerpen aan de branch en bound methode zal ik tonen dat de systeemgrootte waarvoor het probleem oplosbaar is toeneemt tot $N = 50$. Voor een systeem met deze grootte betekent dit dat we enkel 10 componenten moeten knippen vermits een branch en bound enumeratie mogelijk is voor 40 componenten. In de limiet voor $N \rightarrow \infty$ blijkt dat bij het knippen van de 20% sterkste componenten van de quasi-optimale hyperkubus precursor meer

³Na exploratie van knoop $m+1$ via de eerste vertakking vanuit knoop m wordt de tweede vertakking bekeken door de $m+1$ -de component van B van teken te veranderen.

⁴Dit is de minimale stabiliteit van de voorlopig gevonden binaire vektor.

⁵De gebruikte gelijkheid is hier de gelijkheid gebruikt in FORTRAN. Dit wil zeggen: de waarde van κ wordt toegekend aan κ_b -schatting.

⁶De boom wordt verder geëxploreerd met een onveranderde κ_b -schatting.



Figuur 5.2: De fractie $f_x(\alpha)$ met $x = 0.2$, voor de quasi-optimale hyperkubus precursor bepaald door de potentiaal $V_{qo}(\lambda)$

dan 99% van de geknipte synapsen correct hun MSB component voorspellen (Zie figuur 5.2). Dus voor het relatief kleine systeem van $N = 50$ zullen er bijna geen foute componenten aanwezig zijn in die 10 geknipte koppelingen. Dit heeft tot gevolg dat er een exacte oplossing mag verwacht worden na het bepalen van de overige 40 componenten met branch en bound. Merk op dat deze strategie niet uitbreidbaar is naar waarden van N die veel groter zijn dan $N = 50$ omdat er dan veel meer dan 20% van de componenten dienen geknipt te worden, wat onvermijdelijk een aantal foute voorspellingen zal opleveren zodat een exacte oplossing onbereikbaar wordt⁷ Merk op dat de numerieke berekening van de hyperkubus precursor heel snel kan gebeuren met een gradient descent methode. De oplossingen voor $N = 50$ kunnen dan ook bekomen worden zonder een wezenlijke toename van computertijd in vergelijking met het toepassen van branch en bound op 40 componenten.

Uit vorige sectie is het onmiddellijk duidelijk dat de branch en bound methode sneller kan gemaakt worden als men de κ_b -schatting verbetert. Immers,

⁷Natuurlijk kan men de resultaten voor $N = 60, 75$ en 100 uit het vorige hoofdstuk verbeteren met de methode van branch en bound. In het geval van $N = 60$ moeten we bijvoorbeeld maar 20 componenten knippen (i. p. v 35) vermits de overige 40 kunnen bepaald worden met het branch en bound algoritme. Men kan dan een zeer goede benadering voor de MSB verwachten. Of de oplossing voor deze grotere systemen ook exact is zal met toenemende waarde van N onwaarschijnlijker worden.

5. Optimale leerstrategie voor het binair perceptron

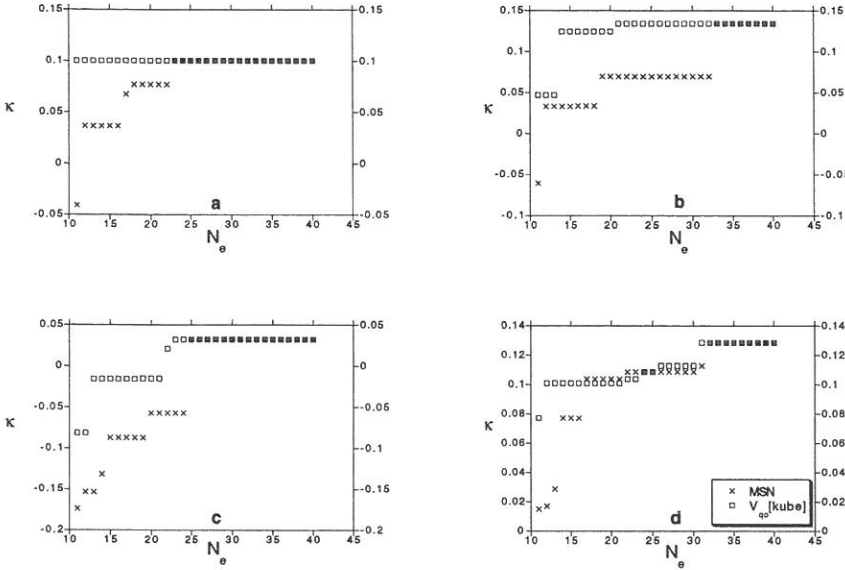
een betere (grotere) κ_b -schatting impliceert dat er meer vertakkingen zullen weggesneden worden in dewelke het algoritme dus geen computertijd dient te verbruiken. In deze leerstrategie zal ik de κ_b schatting bepalen door een kleine fractie van de kleinste componenten van de hyperkubus precursor te enumereren met de Gray-code methode en de rest te knippen. Op deze manier kan natuurlijk de κ_b -schatting verbeterd worden door steeds meer en meer koppelingen te enumereren. Natuurlijk moet men een prijs betalen voor een betere κ_b -schatting. Door meer en meer componenten te willen enumereren zal de tijd nodig om de κ_b -schatting te bepalen sterk stijgen waardoor uiteindelijk de tijdswinst bekomen met de branch en bound teniet zal worden gedaan. Om wille van deze reden kan men de schatting voor κ_b niet blijven verbeteren. De schatting die uiteindelijk gekozen is wordt bepaald uit de binaire vektor bekomen door de 10 kleinste componenten van de precursor te enumereren en de anderen te knippen. Het geeft een voldoende grote κ_b -schatting en het neemt niet veel tijd in beslag om het te bekomen.

5.4 Simulaties

Numerieke simulaties zijn uitgevoerd voor twee systeemgrootten $N = 40$ en 50 . Alle simulaties zijn uitgevoerd bij $\alpha = 0.8$, nabij de maximale geheugen-capaciteit $\alpha_c = 0.83$. Een aantal gewichten is voorspeld door het knippen van de hyperkubus precursor componenten, terwijl de anderen bepaald worden met het branch en bound algoritme. In plaats van echter een vast aantal componenten te enumereren⁸ (en dus een vast aantal te knippen), presenteren de simulaties resultaten voor een variërend aantal geënumereerde componenten. Op deze manier tonen de simulaties de evolutie van de minimale stabiliteit als het aantal geënumereerde componenten N_e toeneemt. Als dit getal klein is zal er een groot aantal synapsen moeten voorspeld worden door het knippen van de overeenkomstige componenten van de precursor. Om wille van de grote fractie van componenten dat op deze manier bepaald wordt zullen er onvermijdelijk fouten optreden, maar deze worden geleidelijk aan gecorrigeerd als N_e toeneemt. Op een bepaald moment is het niet meer nuttig van N_e nog verder toe te laten nemen vermits de overgebleven componenten correct voorspeld zijn door het "knip-proces".

⁸In wat volgt zal ik de term "enumereren" ook gebruiken voor het branch en bound algoritme. Dit algoritme voert weliswaar geen echte enumeratie uit, maar het uiteindelijk resultaat is er wel equivalent mee.

5.4. Simulaties



Figuur 5.3: Simulaties bij $\alpha = 0.8$ voor een perceptron met $N=40$. De figuur toont de evolutie van κ als N_e toeneemt van 11 tot 40. De 4 figuren zijn een selectie uit 20 simulaties met figuur 5.3a en 5.3d respectievelijk het beste en slechtste geval. Figuur 5.3b en 5.3c zijn twee typische gevallen. De legende van figuur 5.3d is ook op de anderen van toepassing.

Om wille van het grote aantal enumeraties dat gedaan dient te worden door N_e te variëren is elke simulatie getoond voor een vaste set van patronen. Ten gevolge van het ontbreken van een middeling over de patronen zullen de bekomen minimale stabiliteiten fluctueren rond hun theoretische waarde $\kappa_b = 0.03$. Het hoofddoel van de simulaties is echter te tonen dat de bekomen oplossing exact is. Er is dan ook minder belang gehecht aan de bekomen waarde van κ . Alle simulaties zijn uitgevoerd met de quasi-optimale hyperkubus precursor en met de MSN als vergelijking.

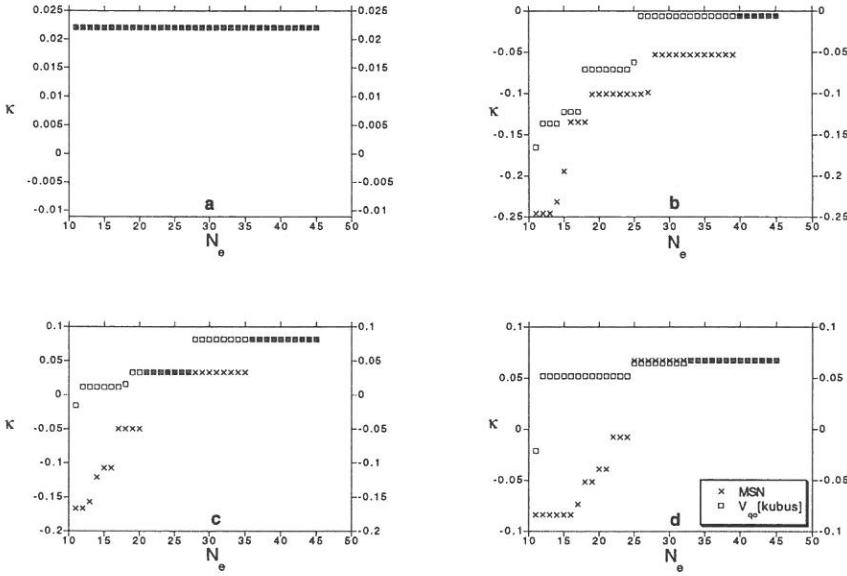
Figuur 5.3 toont de evolutie van κ voor een perceptron met $N = 40$ door N_e toe te laten nemen van $N_e = 11$ tot 40. Op het laatst wordt er dus zeker een exacte oplossing bekomen daar $N_e = 40$ overeenkomt met een volledige

5. Optimale leerstrategie voor het binair perceptron

enumeratie van alle componenten. Vier figuren tonen een selectie uit 20 soortgelijke simulaties telkens uitgevoerd met een andere vaste set van patronen. Figuur 5.3a toont het beste geval uit deze reeks simulaties. De uiteindelijke κ waarde wordt reeds bekomen door enkel 11 componenten van de hyperkubus precursor te enumereren. Figuur 5.3d toont het slechtste geval waar minimum 31 componenten moeten geënumereerd worden om de uiteindelijke minimale stabiliteit te bekomen. Figuur 5.3b en figuur 5.3c tonen twee typische gevallen tussen deze beide voorgaande extremen. Duidelijk blijkt uit alle 4 figuren dat een groot aantal componenten correct voorspeld zijn door de hyperkubus precursor. Dit aantal loopt op van 9 in het slechtste geval, doorheen 22 in een typisch geval, tot 29 in het beste geval. Dit suggereert dat men met zeer grote zekerheid ongeveer 20% van de sterkste componenten van de hyperkubus precursor mag knippen. Dit is in overeenstemming met de theoretische resultaten voor $N \rightarrow \infty$ uit figuur 5.2. We kunnen dus zelfs concluderen dat men het geval voor $N = 40$ exact kan oplossen door enkel de 32 kleinste componenten te enumereren en de grootste 8 te knippen. Het voordeel van de hyperkubus precursor komt duidelijk naar voor in al deze figuren. In de meeste gevallen is de hyperkubus ver vooruit op de MSN in zijn evolutie naar het exacte resultaat.

Figuur 5.4 toont de resultaten voor $N = 50$ waar een volledige enumeratie onmogelijk is. Het getal N_e neemt toe van 11 tot 45, zodat er enkel 5 componenten niet geënumereerd worden. Zodra $N_e > 40$ neemt de vereiste computertijd heel vlug onpraktische waarden aan. Daarom zijn er in dit geval slechts 8 simulaties uitgevoerd. Figuur 5.4a toont weer het beste geval. Figuur 5.4b en figuur 5.4c tonen twee typische gevallen en figuur 5.4d het slechtste geval uit de 8 simulaties. Vergelijking met de vorige gevallen voor $N = 40$ suggereert dat de 5 niet geënumereerde componenten —die de sterkste zijn van de hyperkubus precursor— met zeer grote waarschijnlijkheid correct geknipt zijn zodat de κ -waarde gevonden bij $N_e = 45$ exact is. Als we aannemen dat dit laatste waar is zien we dat deze waarde in alle gevallen bereikt wordt voor $N_e > 35$. Dit betekent dat het knippen van de 10 sterkste componenten van de precursor en het enumereren van de 40 anderen bijna met zekerheid leidt tot een exact resultaat voor $N = 50$. Dit is weerom in overeenstemming met wat men theoretisch verwacht voor $N \rightarrow \infty$.

Het voordeel van de hyperkubus precursor komt ook hier weer duidelijk naar voor in al deze figuren. In vorig hoofdstuk was dit reeds getoond voor een enumeratie van 25 componenten. In deze nieuwe figuren zien we nu echter de evolutie van de minimale stabiliteit als meer en meer fout geknipte componenten gecorrigeerd worden door de enumeratiemethode. In de meeste gevallen is



Figuur 5.4: Simulaties bij $\alpha = 0.8$ voor een perceptron met $N=50$. De figuur toont de evolutie van κ bij een toenemende waarde van N_e . De 4 figuren zijn een selectie uit 8 simulaties met figuur 5.4a en 5.4d respectievelijk het beste en slechtste. Figuur 5.4b en 5.4c zijn twee typische gevallen. De legende van figuur 5.4d is ook op de anderen van toepassing.

de hyperkubus ver vooruit op de MSN in zijn evolutie naar het exacte resultaat. Soms is dit echter niet waar zoals kan gezien worden in figuur 5.3d en 5.4d. Een verbeterde strategie zou dus kunnen zijn van beide precursors te gebruiken, zowel de MSN als de quasi-optimale hyperkubus precursor, om de MSB te vinden. Het verdubbelt de benodigde computertijd maar het geeft een extra garantie dat de MSB correct gevonden wordt.

5. Optimale leerstrategie voor het binair perceptron

Hoofdstuk 6

Conclusie

In deze thesis is het capaciteitsprobleem van het binair perceptron uitvoerig behandeld. Vermits dit probleem een NP-compleet probleem is, is het tot vandaag de dag onmogelijk om het op te lossen voor arbitraire systeemgrootten. Exacte oplossingen zijn mogelijk maar dan moet $N \leq 40$. Voor zulke relatief kleine systemen kan men methodes zoals "Gray-coding" en branch en bound gebruiken om oplossingen te vinden. Deze thesis is voornamelijk een zoektocht naar methodes om een zeker aantal componenten van de MSB exact te kunnen voorspellen. Zo kunnen dan systemen groter dan $N = 40$ beschouwd worden. (Indien men $N - 40$ componenten zou kunnen voorspellen, kan men de overige 40 synapsen onderwerpen aan een branch en bound algoritme om tot een exacte oplossing te komen.)

De vraag die dan naar voor komt is hoe we een zeker aantal MSB componenten correct kunnen voorspellen? Vermits er efficiënte algoritmen bestaan voor het capaciteitsprobleem met continue koppelingen kan men trachten uit deze koppelingen de gezochte binaire koppelingen te bepalen. Dit idee om een continu perceptron te gebruiken als voorbode, of precursor, op het vinden van een binair perceptron met maximale stabiliteit is in deze thesis uitvoerig behandeld. Ik heb onderzocht hoeveel informatie over de MSB kan bekomen worden uit een perceptron met continue koppelingen door het knippen van zijn componenten. In het bijzonder heb ik onderzocht hoeveel componenten bij knippen correct de overeenkomstige component van de MSB voorspellen. Tevens heb ik getoond hoe men deze componenten kan identificeren.

Ik heb verschillende precursors beschouwd die allen bepaald worden door het minimum van een kostfunctie van de volgende vorm: $E(\mathbf{J}) = \sum_{\mu} V(\lambda_{\mu})$. De precursor leerregel is volledig bepaald door specificatie van de potentiaal

6. Conclusie

$V(\lambda)$ en het minimalistiedomein van de kostfunctie. Allereerst heb ik me beperkt tot precursors met de hypersfeer $\mathbf{J}^2 = N$ als minimalisatiedomein. Hiertoe behoort ook de MSN. Ik heb in dit geval een uitdrukking afgeleid voor de maximale fractie van correct geknipte componenten van de MSB, bereikbaar met een precursor geconstrueerd met de beschouwde klasse van kostfuncties op de hypersfeer. Verder is er een potentiaal $V_{qo}(\lambda)$ voorgesteld die deze maximale fractie zeer goed benaderd. Ik heb hem de quasi-optimale potentiaal genoemd.

Door verandering van het minimalisatiedomein van de hypersfeer naar de ingeschreven hyperkubus $|J_i| \leq 1$ bleek de precursor geconstrueerd met dezelfde quasi-optimale potentiaal $V_{qo}(\lambda)$ nog fel te verbeteren. Deze verbetering, in vergelijking met andere precursors zoals de MSN, is bekomen door zowel de hyperkubus constraint als het gebruik van de quasi optimale potentiaal. De analytische en numerieke resultaten wijzen erop dat de hyperkubus precursor het capaciteitsprobleem kan vereenvoudigen. Voor kleine waarden van α kan men ongeveer 60% van de binaire componenten bekomen uit de precursor. Voor grotere waarden van α zijn ongeveer 40% van de binaire componenten correct voorspeld. De mogelijkheid tot het voorspellen van bepaalde componenten vervangt het oorspronkelijke probleem door een eenvoudiger probleem van kleinere grootte. De numerieke simulaties die ik uitgevoerd heb suggereren dat zelfs grotere fracties van de hyperkubus precursor mogen geknipt worden. Dit zal slechts resulteren in een kleine afname van de minimale stabiliteit. Dit numeriek resultaat suggereert dat de hyperkubus precursor die binaire componenten correct voorspeld die essentieel zijn voor het bekomen van een grote minimale stabiliteit. Bovendien zijn de componenten die fout voorspeld worden niet cruciaal voor het bekomen van een grote minimale stabiliteit.

In een laatste hoofdstuk heb ik, door de quasi-optimale hyperkubus precursor in combinatie met een branch en bound methode te gebruiken, oplossingen bekomen voor het binair capaciteitsprobleem tot $N = 50$. Voor grotere systeemgrootten is het zeker mogelijk van met deze methode een goede benadering te bekomen voor de MSB vektor, maar exacte oplossingen zijn vrijwel onmogelijk.

Bijlage A

Replica symmetrie berekening voor $P(j, b)$

A.1 Introductie van replicas

De grootheid (2.7) kan met de introductie van n replicas geschreven worden als:

$$P(j, b) = \lim_{n \rightarrow 0} \left\langle \int_{-\infty}^{+\infty} d\mu(\mathbf{J}) d\mu(\mathbf{B}) \delta(J_1 - j) \delta(B_1 - b) \exp[-\beta \sum_{\mu} (V(\lambda_{\mu}) + \Phi(\Lambda_{\mu}))] \cdot \left[\int_{-\infty}^{+\infty} d\mu(\mathbf{J}) d\mu(\mathbf{B}) \exp[-\beta \sum_{\mu} (V(\lambda_{\mu}) + \Phi(\Lambda_{\mu}))] \right]^{n-1} \right\rangle.$$

Hierbij is $d\mu(\mathbf{J}) = \prod_i p(J_i) dJ_i$ en $d\mu(\mathbf{B}) = \prod_i p(B_i) dB_i$. Verder is de potentiaal V en $p(J_i)$ alsnog willekeurig en de potentiaal Φ en $p(B_i)$ bepalen de MSB leerregel en zijn gegeven door (2.5). Als we de n replicas een index $a = 1, \dots, n$ geven kan men noteren:

$$P(j, b) = \lim_{n \rightarrow 0} \int_{-\infty}^{+\infty} \prod_a^n (d\mu(\mathbf{J}^a) d\mu(\mathbf{B}^a)) \delta(J_1^{a=1} - j) \delta(B_1^{a=1} - b) \cdot \prod_{\mu} \left\langle \exp[-\beta \sum_a (V(\lambda_{\mu}^a) + \Phi(\Lambda_{\mu}^a))] \right\rangle \quad (\text{A.1})$$

A. Replica symmetrie berekening voor $P(j, b)$

Hier hebben we tevens gebruik gemaakt van de onafhankelijkheid van de distributies van de patronen η^μ door het produkt over μ buiten het gemiddelde te brengen. Vermits alle patronen ook identiek verdeeld zijn kunnen we tevens $\Pi_\mu < \dots >$ vervangen door $(< \dots >)^p$, waarbij het gemiddelde uitgevoerd wordt voor een willekeurige en vaste μ -waarde. De μ -index in de stabiliteiten λ_μ^a en Λ_μ^a wordt dus overbodig en zal in wat volgt weggelaten worden. Merk op dat de beide potentialen tegelijk gemiddeld worden over het patroon η^μ . Dit verzekert dat beide leerregels getraind worden met dezelfde set van patronen. Het is net dit soort middeling wat de koppeling tussen de twee systemen zal veroorzaken. Met behulp van

$$P(\bar{\lambda}_1, \dots, \bar{\lambda}_n, \bar{\Lambda}_1, \dots, \bar{\Lambda}_n) = \left\langle \prod_a \delta(\bar{\lambda}_a - \lambda^a) \delta(\bar{\Lambda}_a - \Lambda^a) \right\rangle, \quad (\text{A.2})$$

kan het gemiddelde in (A.1) geschreven worden als:

$$\int_{-\infty}^{+\infty} \prod_a (d\bar{\lambda}_a d\bar{\Lambda}_a) P(\bar{\lambda}_1, \dots, \bar{\Lambda}_n) \exp[-\beta \sum_a (V(\bar{\lambda}_a) + \Phi(\bar{\Lambda}_a))]. \quad (\text{A.3})$$

Men kan $P(\bar{\lambda}_1, \dots, \bar{\Lambda}_n)$ berekenen met behulp van zijn Fouriergetransformeerde, de zogenaamde karakteristieke functie. Volgens (E.8) wordt deze karakteristieke functie dan gegeven door:

$$P(\bar{x}_1, \dots, \bar{x}_n, \bar{X}_1, \dots, \bar{X}_n) = \langle \exp[-i \sum_a (\lambda^a \bar{x}_a + \Lambda^a \bar{X}_a)] \rangle = \langle \exp[-i Y] \rangle. \quad (\text{A.4})$$

Met $Y = \frac{1}{\sqrt{N}} \sum_j \sum_a J_j^a \eta_j \bar{x}_a + \frac{1}{\sqrt{N}} \sum_j \sum_a B_j^a \eta_j \bar{X}_a$. Voor zeer grote waarde van N is Y wegens het centraal limiet theorema Gaussisch verdeeld met $\langle Y \rangle = 0$ en variantie σ^2 gegeven door:

$$\frac{1}{N} \sum_j \sum_{ab} J_j^a J_j^b \bar{x}_a \bar{x}_b + \frac{1}{N} \sum_j \sum_{ab} B_j^a B_j^b \bar{X}_a \bar{X}_b + \frac{2}{N} \sum_j \sum_{ab} J_j^a B_j^b \bar{x}_a \bar{X}_b. \quad (\text{A.5})$$

Hieruit volgt dan onmiddellijk dat $P(\bar{x}_1, \dots, \bar{X}_n) = \langle \exp[-i Y] \rangle = \exp[-\frac{1}{2} \sigma^2]$. Invers Fouriertransformeren (E.9) van de karakteristieke functie levert de distributie $P(\bar{\lambda}_1, \dots, \bar{\Lambda}_n)$ en deze invullen in (A.3) levert dan voor het gemiddelde in (A.1):

A.1. Introductie van replicas

$$\begin{aligned}
 & \int_{-\infty}^{+\infty} \prod_a \left(\frac{dx_a dX_a d\lambda_a d\Lambda_a}{(2\pi)^2} \right) \exp[-\beta \sum_a (V(\lambda_a) + \Phi(\Lambda_a))] \cdot \\
 & \cdot \exp[i \sum_a (x_a \lambda_a + X_a \Lambda_a)] \exp[-\sum_{ab} r_{ab} x_a X_b] \cdot \\
 & \cdot \exp[-\frac{1}{2} \sum_a (x_a)^2 f_a - \frac{1}{2} \sum_a (X_a)^2] \cdot \\
 & \cdot \exp[-\sum_{a<b} x_a x_b q_{ab} - \sum_{a<b} X_a X_b Q_{ab}], \tag{A.6}
 \end{aligned}$$

waar we de streepjes boven sommige variabelen zoals $\overline{x_a}$ weggelaten hebben en tevens de volgende ordeparameters ingevoerd hebben:

$$\begin{aligned}
 f_a &= \frac{1}{N} \sum_j (J_j^a)^2; \quad r_{ab} = \frac{1}{N} \sum_j J_j^a B_j^b \\
 q_{ab} &= \frac{1}{N} \sum_j J_j^a J_j^b; \quad Q_{ab} = \frac{1}{N} \sum_j B_j^a B_j^b \tag{A.7}
 \end{aligned}$$

Met het invoeren van Dirac-delta functies kan men nu de uitdrukking (A.1) voor $P(j, b)$ schrijven als:

$$\begin{aligned}
 & \lim_{n \rightarrow 0} \int_{-\infty}^{+\infty} \prod_a (d\mu(J^a) d\mu(B^a)) \delta(J_1^{a=1} - j) \delta(B_1^{a=1} - b) \cdot \\
 & \cdot \int_{-\infty}^{+\infty} \prod_{ab} dr_{ab} \delta(r_{ab} - \frac{1}{N} \sum_j J_j^a B_j^b) \cdot \\
 & \cdot \int_{-\infty}^{+\infty} \prod_a df_a \delta(f_a - \frac{1}{N} \sum_j (J_j^a)^2) \cdot \\
 & \cdot \int_{-\infty}^{+\infty} \prod_{a<b} dq_{ab} dQ_{ab} \delta(q_{ab} - \frac{1}{N} \sum_j J_j^a J_j^b) \delta(Q_{ab} - \frac{1}{N} \sum_j B_j^a B_j^b) \cdot \\
 & \cdot [(A.6)]^p. \tag{A.8}
 \end{aligned}$$

Als we Dirac-delta functies schrijven m. b. v. hun integraalrepresentatie (E.21) krijgt elke ordeparameter zijn toegevoegde. Vb: f_a en zijn toegevoegde $\hat{f}_a$. Met

A. Replica symmetrie berekening voor $P(j, b)$

de volgende reeks substituties:

$$\begin{aligned}\hat{f}'_a &= \frac{i}{N} \hat{f}_a; & \hat{q}'_{ab} &= \frac{-i}{N} \hat{q}_{ab} \\ \hat{Q}'_{ab} &= \frac{-i}{N} \hat{Q}_{ab}; & \hat{r}'_{ab} &= \frac{-i}{N} \hat{r}_{ab},\end{aligned}$$

bekomen we voor $P(j, b)$ (waarbij we de accenten bij de nieuwe variabelen weggelaten hebben):

$$\begin{aligned}& \lim_{n \rightarrow 0} \int d[f_a, \hat{f}_a, q_{ab}, \hat{q}_{ab}, Q_{ab}, \hat{Q}_{ab}, r_{ab}, \hat{r}_{ab}] \cdot \\& \cdot \exp \left[N \left(\sum_a f_a \hat{f}_a - \sum_{a < b} q_{ab} \hat{q}_{ab} - \sum_{a < b} Q_{ab} \hat{Q}_{ab} - \right. \right. \\& \left. \left. - \sum_{ab} r_{ab} \hat{r}_{ab} + \alpha \ln G_o + \frac{N-1}{N} \ln G_1 \right) \right] \cdot \\& \cdot \int_{-\infty}^{+\infty} \prod_a (d\mu(J_a) d\mu(B_a)) \delta(J^{a=1} - j) \delta(B^{a=1} - b) \exp \left[- \sum_a \hat{f}_a (J_a)^2 + \right. \\& \left. + \sum_{a < b} \hat{q}_{ab} J_a J_b + \sum_{a < b} \hat{Q}_{ab} B_a B_b + \sum_{ab} \hat{r}_{ab} J_a B_b \right] \quad (\text{A.9})\end{aligned}$$

waarbij $d\mu(J_a) = p(J_a) dJ_a$, $d\mu(B_a) = p(B_a) dB_a$ en waarbij $d[f_a, \hat{f}_a, \dots, \hat{r}_{ab}]$ symbool staat voor alle integratiematen. Verder is $\alpha = p/N$, de geheugen-bezetting en wordt $G_o = G_o(f_a, q_{ab}, Q_{ab}, r_{ab})$ gegeven door (A.6) en G_1 door:

$$\begin{aligned}G_1 &= G_1(\hat{f}_a, \hat{q}_{ab}, \hat{Q}_{ab}, \hat{r}_{ab}) = \\&= \int_{-\infty}^{+\infty} \prod_a d\mu(J_a) d\mu(B_a) \exp \left[- \sum_a \hat{f}_a (J_a)^2 + \right. \\& \left. + \sum_{a < b} \hat{q}_{ab} J_a J_b + \sum_{a < b} \hat{Q}_{ab} B_a B_b + \sum_{ab} \hat{r}_{ab} J_a B_b \right] \quad (\text{A.10})\end{aligned}$$

A.2 De RS-ansatz

Voor $N \rightarrow \infty$ wordt de integraal over de ordeparameters (en hun toegevoegden) in (A.9) enkel bepaald door het zadelpunt van de coefficient van N in de

A.2. De RS-ansatz

exponent. We kunnen dan de laatste factor die niet afhangt van N buiten deze integraal brengen waarbij de waarde in het zadelpunt van de ordeparameters en hun toegevoegden ingevuld dienen te worden. Hierbij gebruiken we dus de formule:

$$\lim_{N \rightarrow \infty} \int_C dz e^{Nf(z)} g(a, z) = g(a, z_o) \lim_{N \rightarrow \infty} \int_C dz e^{Nf(z)}, \quad (\text{A.11})$$

waarbij z_o het zadelpunt is van $f(z)$ en dus bepaald is door $f'(z_o) = 0$. Als we dit toepassen op (A.9) en tevens de RS-ansatz maken voor het zadelpunt:

$$\begin{aligned} f_a &= f & \forall a \\ \hat{f}_a &= \hat{f} & \forall a \\ q_{ab} &= q & \forall a < b \\ \hat{q}_{ab} &= \hat{q} & \forall a < b \\ Q_{ab} &= Q & \forall a < b \\ \hat{Q}_{ab} &= \hat{Q} & \forall a < b \\ r_{ab} &= r & \forall a, b \\ \hat{r}_{ab} &= \hat{r} & \forall a, b \end{aligned} \quad (\text{A.12})$$

vinden we:

$$\begin{aligned} P(j, b) &= \lim_{n \rightarrow 0} \int_{-\infty}^{+\infty} \prod_a d\mu(J_a) d\mu(B_a) \delta(J_{a=1} - j) \delta(B_{a=1} - b) \\ &\cdot \exp \left[-\hat{f} \sum_a (J_a)^2 + \hat{q} \sum_{a < b} J_a J_b + \right. \\ &\quad \left. + \hat{Q} \sum_{a < b} B_a B_b + \hat{r} \sum_{ab} J_a B_b \right]. \end{aligned} \quad (\text{A.13})$$

Bemerk dat $P(j, b)$ enkel afhangt van de toegevoegden van de ordeparameters, de waarden die deze toegevoegden aannemen worden bepaald door:

$$\begin{aligned} \lim_{n \rightarrow 0} \text{Extr}_{\{q, Q, f, r, \hat{q}, \hat{Q}, \hat{f}, \hat{r}\}} &\left[n f \hat{f} - \frac{n(n-1)}{2} q \hat{q} - \frac{n(n-1)}{2} Q \hat{Q} - \right. \\ &\quad \left. - n^2 r \hat{r} + \alpha \ln G_o(q, Q, f, r) + \ln G_1(\hat{q}, \hat{Q}, \hat{f}, \hat{r}) \right], \end{aligned} \quad (\text{A.14})$$

waarbij G_o en G_1 gegeven zijn door:

A. Replica symmetrie berekening voor $P(j, b)$

$$\begin{aligned}
 G_o = & \int_{-\infty}^{+\infty} \prod_a \left(\frac{dx_a dX_a d\lambda_a d\Lambda_a}{(2\pi)^2} \right) \exp[-\beta \sum_a (V(\lambda_a) + \Phi(\Lambda_a))] \cdot \\
 & \cdot \exp[i \sum_a (x_a \lambda_a + X_a \Lambda_a)] \exp[-r(\sum_a x_a)(\sum_b X_b)] \cdot \\
 & \cdot \exp[-\frac{1}{2}(f - q) \sum_a (x_a)^2 - \frac{1}{2}(1 - Q) \sum_a (X_a)^2] \cdot \\
 & \cdot \exp[-\frac{q}{2}(\sum_a x_a)^2 - \frac{Q}{2}(\sum_a X_a)^2], \tag{A.15}
 \end{aligned}$$

en

$$\begin{aligned}
 G_1 = & \int_{-\infty}^{+\infty} \prod_a d\mu(J_a) d\mu(B_a) \cdot \\
 & \cdot \exp \left[-\frac{1}{2}(2\hat{f} + \hat{q}) \sum_a (J_a)^2 - \frac{\hat{Q}}{2} \sum_a (B_a)^2 + \right. \\
 & \left. + \frac{\hat{q}}{2}(\sum_a J_a)^2 + \frac{\hat{Q}}{2}(\sum_a B_a)^2 + \hat{r}(\sum_a J_a)(\sum_b B_b) \right] \tag{A.16}
 \end{aligned}$$

A.3 De limiet $n \rightarrow 0$

Om de limiet voor $n \rightarrow 0$ te nemen in (A.14) moet men eerst de uitdrukkingen voor G_o en voor G_1 herschrijven zodat de n -afhankelijkheid expliciet naar voor komt. Hiertoe moet men de koppelingen tussen de verschillende replica indices verwijderen. Ik zal als voorbeeld de uitdrukking voor G_o herschrijven, G_1 is volledig analoog. Beschouw de term $\exp[-r(\sum_a x_a)(\sum_b X_b) - \frac{q}{2}(\sum_a x_a)^2 - \frac{Q}{2}(\sum_a X_a)^2]$ in de uitdrukking (A.15). Deze bevat de koppelingen tussen de replica indices. Door toepassing van achtereenvolgens (E.25) en (E.13) kan men deze term herschrijven tot:

$$\begin{aligned}
 \int_{-\infty}^{+\infty} Dt Du \exp \left[-it\sqrt{1+\gamma} \left(\sqrt{\frac{q}{2}} \sum_a x_a + \sqrt{\frac{Q}{2}} \sum_a X_a \right) \right. \\
 \left. -iu\sqrt{1-\gamma} \left(\sqrt{\frac{q}{2}} \sum_a x_a - \sqrt{\frac{Q}{2}} \sum_a X_a \right) \right], \tag{A.17}
 \end{aligned}$$

waarbij $\gamma = \frac{r}{\sqrt{qQ}}$. Dit invullen in de uitdrukking (A.15) voor G_o levert:

$$\begin{aligned}
 & \int_{-\infty}^{+\infty} Dt Du \left(\int_{-\infty}^{+\infty} \frac{d\lambda dx d\Lambda dX}{(2\pi)^2} \exp[-\beta(V(\lambda) + \Phi(\Lambda))] \cdot \right. \\
 & \cdot \exp \left[i(\lambda x + \Lambda X) - \frac{f-q}{2} x^2 - \frac{1-Q}{2} X^2 - \right. \\
 & - i\sqrt{\frac{q}{2}} x \left(t\sqrt{1+\gamma} + u\sqrt{1-\gamma} \right) - \\
 & \left. \left. - i\sqrt{\frac{Q}{2}} X \left(t\sqrt{1+\gamma} - u\sqrt{1-\gamma} \right) \right] \right)^n. \quad (A.18)
 \end{aligned}$$

De integraties over t en u kunnen met de volgende substitutie vereenvoudigd worden:

$$\begin{aligned}
 \sigma &= \frac{1}{\sqrt{2}}(t\sqrt{1+\gamma} + u\sqrt{1-\gamma}) \\
 \tau &= \frac{1}{\sqrt{2}}(t\sqrt{1+\gamma} - u\sqrt{1-\gamma}). \quad (A.19)
 \end{aligned}$$

* Dit geeft voor de integratiemaat:

$$Du Dt = \frac{d\sigma d\tau}{2\pi\sqrt{1-\gamma^2}} \exp \left[-\frac{1}{2(1-\gamma^2)} (\sigma^2 + \tau^2 - 2\gamma\sigma\tau) \right] \equiv D_\gamma(\sigma, \tau), \quad (A.20)$$

waarin de factor $\frac{1}{\sqrt{1-\gamma^2}}$ de determinant is van de Jacobiaanse matrix van de transformatie. De substitutie uitvoeren in (A.18) en de integraties uitvoeren naar x en X met behulp van (E.10) levert dan:

$$\begin{aligned}
 G_o &= \int_{-\infty}^{+\infty} D_\gamma(\sigma, \tau) \left(\int_{-\infty}^{+\infty} \frac{d\lambda}{\sqrt{2\pi(f-q)}} \frac{d\Lambda}{\sqrt{2\pi(1-Q)}} \cdot \right. \\
 & \cdot \exp \left[-\beta V(\lambda) - \frac{1}{2(f-q)} (\lambda - \sqrt{q}\sigma)^2 \right] \cdot \\
 & \cdot \exp \left[-\beta \Phi(\Lambda) - \frac{1}{2(1-Q)} (\Lambda - \sqrt{Q}\tau)^2 \right] \right)^n. \quad (A.21)
 \end{aligned}$$

Een analoge manier van werken levert voor G_1 in (A.16):

$$G_1 = \int_{-\infty}^{+\infty} D_\gamma(\varphi, \psi) \left(\int_{-\infty}^{+\infty} d\mu(J) d\mu(B). \quad (A.22)$$

A. Replica symmetrie berekening voor $P(j, b)$

$$\cdot \exp \left[-\frac{1}{2}(2\hat{f} + \hat{q})J^2 - \frac{\hat{Q}}{2} + \sqrt{\hat{q}}J\varphi + \sqrt{\hat{Q}}B\psi \right]^n, \quad (\text{A.23})$$

waarbij $\hat{\gamma} = \frac{\hat{r}}{\sqrt{\hat{q}\hat{Q}}}$.

Men is nu klaar om de limiet voor $n \rightarrow 0$ te nemen in vergelijking (A.14). Men vindt tot op eerste orde in n :

$$\begin{aligned} & \text{Extr}_{\{q, Q, f, r, \hat{q}, \hat{Q}, \hat{f}, \hat{r}\}} \left\{ f\hat{f} + \frac{1}{2}q\hat{q} + \right. \\ & + \alpha \int_{-\infty}^{+\infty} D\sigma \ln \left[\int_{-\infty}^{+\infty} \frac{d\lambda}{\sqrt{2\pi(f-q)}} e^{-\beta V(\lambda) - \frac{1}{2(f-q)}(\lambda - \sqrt{q}\sigma)^2} \right] + \\ & + \int_{-\infty}^{+\infty} D\varphi \ln \left[\int_{-\infty}^{+\infty} d\mu(J) \exp \left[-\frac{1}{2}(2\hat{f} + \hat{q})J^2 + \sqrt{\hat{q}}J\varphi \right] \right] + \\ & + \frac{1}{2}Q\hat{Q} + \\ & + \alpha \int_{-\infty}^{+\infty} D\tau \ln \left[\int_{-\infty}^{+\infty} \frac{d\Lambda}{\sqrt{2\pi(1-Q)}} e^{-\beta \Phi(\Lambda) - \frac{1}{2(1-Q)}(\Lambda - \sqrt{Q}\tau)^2} \right] + \\ & \left. + \int_{-\infty}^{+\infty} D\psi \ln \left[\int_{-\infty}^{+\infty} d\mu(B) \exp \left[-\frac{\hat{Q}}{2} + \sqrt{\hat{Q}}B\psi \right] \right] \right\}. \quad (\text{A.24}) \end{aligned}$$

Waar we tevens gebruik gemaakt hebben van de eigenschap (E.19) van de twee dimensionale Gaussische integratiemaat.

Hieruit kunnen we dan heel eenvoudig de zadelpuntvergelijkingen berekenen voor $f, \hat{f}, q, \hat{q}, Q$ en $\hat{Q}$ maar blijkbaar niet voor r en $\hat{r}$ (of γ en $\hat{\gamma}$). Dit komt omdat we in de limiet voor $n \rightarrow 0$ enkel termen meegenomen tot eerste orde in n . Inspectie van (A.14) suggereert dat de zadelpuntvergelijkingen voor r en $\hat{r}$ moeten gezocht worden voor $n \rightarrow 0$ in termen van orde n^2 . We zullen deze nauwkeurigere berekening even uitstellen en voorlopig enkel kijken naar (A.24). We zien dat de zadelpuntvergelijkingen voor $f, \hat{f}, q, \hat{q}$ en $Q, \hat{Q}$ juist die zijn van respectievelijk het precursor systeem $\mathbf{J}$ en het MSB systeem $\mathbf{B}$, zoals ze gewoonlijk optreden [9, 10, 20]. Uit (A.24) zullen we dus de zadelpuntvergelijkingen kunnen afleiden voor het precursor systeem bepaalt door de potentiaal

A.4. De zadelpuntvergelijkingen voor het precursor systeem

V en de distributie $p(J_i)$. Tevens kunnen we ook de zadelpuntvergelijkingen voor Q en $\hat{Q}$ van het MSB systeem bepalen. Hierbij zijn de potentiaal Φ en de distributie $p(B_i)$ gegeven door (2.5).

A.4 De zadelpuntvergelijkingen voor het precursor systeem

Als we $f\hat{f} + \frac{1}{2}q\hat{q}$ schrijven als

$$\frac{1}{2}(f-q)(2\hat{f} + \hat{q}) + \frac{q}{2}(2\hat{f} + \hat{q}) - \frac{1}{2}\hat{q}(f-q), \quad (\text{A.25})$$

dan kan men de zadelpuntvergelijkingen eenvoudig bekomen door af te leiden naar $q, (f-q), \hat{q}$ en $(2\hat{f} + \hat{q})$. Zo bekomen we:

$$\frac{1}{2}(2\hat{f} + \hat{q}) + \alpha \int_{-\infty}^{+\infty} D\sigma \frac{\int_{-\infty}^{+\infty} d\lambda e^{-\beta V(\lambda) - \frac{1}{2(f-q)}(\lambda - \sqrt{q}\sigma)^2} \cdot \frac{(\lambda - \sqrt{q}\sigma)\sigma}{2(f-q)\sqrt{q}}}{\int_{-\infty}^{+\infty} d\lambda e^{-\beta V(\lambda) - \frac{1}{2(f-q)}(\lambda - \sqrt{q}\sigma)^2}} = 0 \quad (\text{A.26})$$

$$\hat{f} - \frac{\alpha}{2(f-q)} + \alpha \int_{-\infty}^{+\infty} D\sigma \frac{\int_{-\infty}^{+\infty} d\lambda e^{-\beta V(\lambda) - \frac{1}{2(f-q)}(\lambda - \sqrt{q}\sigma)^2} \cdot \frac{(\lambda - \sqrt{q}\sigma)^2}{2(f-q)^2}}{\int_{-\infty}^{+\infty} d\lambda e^{-\beta V(\lambda) - \frac{1}{2(f-q)}(\lambda - \sqrt{q}\sigma)^2}} = 0 \quad (\text{A.27})$$

$$-\frac{1}{2}(f-q) + \int_{-\infty}^{+\infty} D\varphi \frac{\int_{-\infty}^{+\infty} d\mu(J) e^{-\frac{1}{2}(2\hat{f} + \hat{q})J^2 + \sqrt{\hat{q}}J\varphi} \cdot \frac{J\varphi}{2\sqrt{\hat{q}}}}{\int_{-\infty}^{+\infty} d\mu(J) e^{-\frac{1}{2}(2\hat{f} + \hat{q})J^2 + \sqrt{\hat{q}}J\varphi}} = 0 \quad (\text{A.28})$$

$$\frac{1}{2}f + \int_{-\infty}^{+\infty} D\varphi \frac{\int_{-\infty}^{+\infty} d\mu(J) e^{-\frac{1}{2}(2\hat{f} + \hat{q})J^2 + \sqrt{\hat{q}}J\varphi} \cdot \left(-\frac{1}{2}J^2\right)}{\int_{-\infty}^{+\infty} d\mu(J) e^{-\frac{1}{2}(2\hat{f} + \hat{q})J^2 + \sqrt{\hat{q}}J\varphi}} = 0 \quad (\text{A.29})$$

Deze zadelpuntvergelijkingen geven bij een bepaalde keuze van de potentiaal V en de distributie $p(J_i)$ informatie over het systeem via de ordeparameters. Indien enkel informatie gewenst is over de perceptrons die de energie minimaliseren moet de limiet voor $\beta \rightarrow \infty$ genomen worden. Met de aanname dat de kostfunctie bepaald door de potentiaal V een uniek minimum heeft geldt

A. Replica symmetrie berekening voor $P(j, b)$

er dat voor $\beta \rightarrow \infty, q \rightarrow f$ en zo kunnen we dan de zadelpuntvergelijkingen makkelijk herschrijven. Als men $\beta(f - q) = x$ stelt dan volgt er uit de methode van Laplace voor de twee zadelpuntvergelijkingen die de potentiaal V bevatten in de limiet voor $\beta \rightarrow \infty$:

$$(2\hat{f} + \hat{q}) + \frac{\alpha}{(f - q)\sqrt{f}} \int_{-\infty}^{+\infty} D\sigma \sigma \left[\lambda_o(\sqrt{f}\sigma, x) - \sqrt{f}\sigma \right] = 0 \quad (\text{A.30})$$

$$\hat{f} - \frac{\alpha}{2(f - q)} + \frac{\alpha}{2(f - q)^2} \int_{-\infty}^{+\infty} D\sigma \left[\lambda_o(\sqrt{f}\sigma, x) - \sqrt{f}\sigma \right]^2 = 0 \quad (\text{A.31})$$

Hierin is λ_o gegeven door:

$$\lambda_o(\sqrt{f}\sigma, x) = \text{Arg Min}_{\lambda} \left[V(\lambda) + \frac{(\lambda - \sqrt{f}\sigma)^2}{2x} \right] \quad (\text{A.32})$$

De vergelijkingen (A.28) en (A.29) zijn in dit stadium niet vereenvoudigbaar omdat de distributie $p(J)$ onbekend is.

A.5 De zadelpuntvergelijkingen voor het MSB systeem

Vermits de potentiaal Φ en $p(B_i)$ gekend zijn kan men in dit geval de zadelpuntvergelijkingen volledig berekenen in de limiet voor $\beta \rightarrow \infty$. Als we in uitdrukking (A.24) enkel rekening houden met de termen die betrekking hebben op het MSB systeem dan vinden we in de limiet voor $\beta \rightarrow \infty$ de volgende vergelijking die moet geëxtremeerd worden naar Q en $\hat{Q}$:

$$\begin{aligned} \text{Extr}_{\{Q, \hat{Q}\}} \left[\frac{Q\hat{Q}}{2} + \alpha \int_{-\infty}^{+\infty} Dz \ln H \left(\frac{\kappa_b - \sqrt{Q}z}{\sqrt{1 - Q}} \right) + \right. \\ \left. + \int_{-\infty}^{+\infty} Dz \ln \left[e^{-\frac{\hat{Q}}{2}} 2 \cosh \left(\sqrt{\hat{Q}}z \right) \right] \right] \end{aligned} \quad (\text{A.33})$$

Hierbij is rekening gehouden met het feit dat $Q \nearrow 1$ als $\beta \rightarrow \infty$ [20]. (Zie ook hoofdstuk 2, sectie 2.4.2.) Deze uitdrukking afleiden naar Q geeft:

A.5. De zadelpuntvergelijkingen voor het MSB systeem

$$\frac{\hat{Q}}{2} + \frac{\alpha}{\sqrt{2\pi}(1-Q)} \int_{-\infty}^{+\infty} Dz \frac{e^{-\frac{(\kappa_b - \sqrt{Q}z)^2}{2(1-Q)}}}{H\left(\frac{\kappa_b - \sqrt{Q}z}{\sqrt{1-Q}}\right)} \cdot \left[\frac{z\sqrt{1-Q}}{2\sqrt{Q}} - \frac{\kappa_b - \sqrt{Q}z}{2\sqrt{1-Q}} \right] = 0 \quad (\text{A.34})$$

Deze vergelijking bevat twee integralen. Uitwerken van de integraal die de factor $\frac{z\sqrt{1-Q}}{2\sqrt{Q}}$ bevat met behulp van (E.15) en terug invullen in vorige vergelijking levert dan:

$$\hat{Q} = \frac{\alpha}{2\pi(1-Q)} \int_{-\infty}^{+\infty} Dz \frac{e^{-\frac{(\kappa_b - \sqrt{Q}z)^2}{(1-Q)}}}{H\left(\frac{\kappa_b - \sqrt{Q}z}{\sqrt{1-Q}}\right)^2} \quad (\text{A.35})$$

Tenslotte leiden we (A.33) af naar $\hat{Q}$ en vinden:

$$1 - Q = \int_{-\infty}^{+\infty} Dz \tanh(\sqrt{\hat{Q}}z) \frac{z}{\sqrt{\hat{Q}}} \quad (\text{A.36})$$

Weer gebruik maken van (E.15) levert dan:

$$1 - Q = \int_{-\infty}^{+\infty} Dz \frac{1}{\cosh^2(\sqrt{\hat{Q}}z)} \quad (\text{A.37})$$

Oplossen van de zadelpuntvergelijkingen (A.35) en (A.37) levert echter niet de ordeparameters voor de MSB vektor. Om de ordeparameters te bekomen voor het maximaal stabiel binair perceptron moeten deze twee zadelpuntvergelijkingen opgelost worden tesamen met de "entropie is nul voorwaarde" [20]. (Zie ook hoofdstuk 2, sectie 2.4.2.) Deze voorwaarde wordt gegeven door:

$$\frac{Q\hat{Q}}{2} + \alpha \int_{-\infty}^{+\infty} Dz \ln H\left(\frac{\kappa_b + \sqrt{Q}z}{\sqrt{1-Q}}\right) + \int_{-\infty}^{+\infty} Dz \ln \left[e^{-\frac{\hat{Q}}{2}} 2 \cosh(\sqrt{\hat{Q}}z) \right] = 0, \quad (\text{A.38})$$

en volgt onmiddellijk uit (A.33). De drie termen zijn immers niets anders zijn dan de entropie bij $E = 0$.

A.6 De zadelpuntvergelijkingen voor r en $\hat{r}$

We zullen nu de vergelijkingen voor r en $\hat{r}$ afleiden. We volgen hierbij de werkwijze van Wong [34] en bepalen de zadelpuntvergelijkingen voor r en $\hat{r}$ vóór de RS-ansatz en dus voor dat $n \rightarrow 0$. De RS-ansatz en de limiet voor $n \rightarrow 0$ nemen we dan in de zadelpuntvergelijkingen zelf. We keren terug naar (A.9). Als $N \rightarrow \infty$ wordt het zadelpunt dan bepaald door:

$$\begin{aligned} \text{Extr}_{\{q, Q, f, r, \hat{q}, \hat{Q}, \hat{f}, \hat{r}\}} \left[\sum_a f_a \hat{f}_a - \sum_{a < b} q_{ab} \hat{q}_{ab} - \sum_{a < b} Q_{ab} \hat{Q}_{ab} - \right. \\ \left. - \sum_{ab} r_{ab} \hat{r}_{ab} + \alpha \ln G_o + \ln G_1 \right], \end{aligned} \quad (\text{A.39})$$

hierbij zijn G_o en G_1 gegeven door respectievelijk (A.6) en (A.10). Uit (A.39) vinden we dan bijvoorbeeld voor de zadelpuntvergelijking voor r_{11} :

$$\begin{aligned} r_{11} = \frac{1}{G_1} \cdot \int_{-\infty}^{+\infty} \prod_a d\mu(J_a) d\mu(B_a) J_1 B_1 \cdot \\ \cdot \exp \left[- \sum_a \hat{f}_a (J_a)^2 + \sum_{a < b} \hat{q}_{ab} J_a J_b + \right. \\ \left. + \sum_{a < b} \hat{Q}_{ab} B_a B_b + \sum_{ab} \hat{r}_{ab} J_a B_b \right] \end{aligned} \quad (\text{A.40})$$

Invoeren van de RS-ansatz (A.12) en de n -afhankelijkheid expliciet naar voor brengen in de integraal met behulp van (E.25) geeft dan:

$$\begin{aligned} r = \frac{1}{G_1} \int_{-\infty}^{+\infty} Dz Dw \left(\int_{-\infty}^{+\infty} d\mu(J) d\mu(B) \cdot \right. \\ \cdot \exp \left[- \frac{1}{2} (2\hat{f} + \hat{q}) J^2 - \frac{\hat{Q}}{2} + \sqrt{\frac{\hat{q}}{2}} J (z\sqrt{1+\hat{\gamma}} + w\sqrt{1-\hat{\gamma}}) + \right. \\ \left. + \sqrt{\frac{\hat{Q}}{2}} B (z\sqrt{1+\hat{\gamma}} - w\sqrt{1-\hat{\gamma}}) \right] \Big)^{n-1} \cdot \\ \cdot \int_{-\infty}^{+\infty} d\mu(J) d\mu(B) J B \exp[\dots] \end{aligned} \quad (\text{A.41})$$

A.6. De zadelpuntvergelijkingen voor r en $\hat{r}$

De "... " in de laatste exponent beduiden dat het dezelfde term betreft als in de vorige exponent. Met behulp van de substitutie (A.19) kunnen we dan voor de zadelpuntvergelijking voor r in de limiet $n \rightarrow 0$ schrijven:

$$r = \int_{-\infty}^{+\infty} D_{\hat{\gamma}}(\varphi, \psi) \cdot \frac{\int_{-\infty}^{+\infty} d\mu(J) d\mu(B) J B \exp [\mathcal{R}(J, B, \varphi, \psi)]}{\int_{-\infty}^{+\infty} d\mu(J) d\mu(B) \exp [\mathcal{R}(J, B, \varphi, \psi)]}, \quad (\text{A.42})$$

waarbij $\mathcal{R}(J, B, \varphi, \psi)$ gegeven wordt door:

$$\mathcal{R} = -\frac{1}{2}(2\hat{f} + \hat{q})J^2 - \frac{\hat{Q}}{2} + \sqrt{\hat{q}}J\varphi + \sqrt{\hat{Q}}B\psi \quad (\text{A.43})$$

Als we de distributie $p(B)$ voor de MSB invullen vereenvoudigt deze vergelijking zich tot:

$$r = \int_{-\infty}^{+\infty} D_{\hat{\gamma}}(\varphi, \psi) \cdot \tanh(\sqrt{\hat{Q}}\psi) \cdot \frac{\int_{-\infty}^{+\infty} d\mu(J) J \exp [\mathcal{R}(J, \varphi)]}{\int_{-\infty}^{+\infty} d\mu(J) \exp [\mathcal{R}(J, \varphi)]}, \quad (\text{A.44})$$

waarbij $\mathcal{R}(J, \varphi)$ nu gegeven wordt door:

$$\mathcal{R} = -\frac{1}{2}(2\hat{f} + \hat{q})J^2 + \sqrt{\hat{q}}J\varphi \quad (\text{A.45})$$

• Een analoge manier van werken levert voor de zadelpuntvergelijking voor $\hat{r}$:

$$\hat{r} = \alpha \int_{-\infty}^{+\infty} D_{\gamma}(\tau, \sigma) \frac{\int_{-\infty}^{+\infty} \frac{d\lambda d\Lambda}{\sqrt{2\pi(f-q)}\sqrt{2\pi(1-Q)}} e^{[\hat{\mathcal{R}}(\lambda, \Lambda, \tau, \sigma)]} \cdot \frac{\lambda - \sqrt{q}\sigma}{f-q} \cdot \frac{\Lambda - \sqrt{Q}\tau}{1-Q}}{\int_{-\infty}^{+\infty} \frac{d\lambda d\Lambda}{\sqrt{2\pi(f-q)}\sqrt{2\pi(1-Q)}} e^{[\hat{\mathcal{R}}(\lambda, \Lambda, \tau, \sigma)]}}, \quad (\text{A.46})$$

waarbij $\hat{\mathcal{R}}(\lambda, \Lambda, \tau, \sigma)$ gegeven wordt door:

$$\hat{\mathcal{R}} = \exp \left[-\beta[V(\lambda) + \Phi(\Lambda)] - \frac{(\lambda - \sqrt{q}\sigma)^2}{2(f-q)} - \frac{(\Lambda - \sqrt{Q}\tau)^2}{2(1-Q)} \right]. \quad (\text{A.47})$$

Als we in vergelijking (A.46) de uitdrukking voor Φ invullen en $\beta \rightarrow \infty$ laten gaan vinden we:

$$\hat{r} = \alpha \int_{-\infty}^{+\infty} D_{\gamma}(\tau, \sigma) \cdot \frac{\lambda_o - \sqrt{q}\sigma}{f-q} \cdot \frac{1}{\sqrt{2\pi(1-Q)}} \cdot \frac{e^{-\frac{(\kappa_b - \sqrt{Q}\tau)^2}{2(1-Q)}}}{H\left(\frac{\kappa_b - \sqrt{Q}\tau}{\sqrt{1-Q}}\right)} \quad (\text{A.48})$$

Hierbij is λ_o bepaald door (A.32). Om deze zadelpuntvergelijkingen te kunnen oplossen is het noodzakelijk van eerst de ordeparameters van de twee afzonderlijke systemen te bepalen en deze in te vullen in de vergelijkingen (A.44) en (A.48).

A. Replica symmetrie berekening voor $P(j, b)$

A.7 De vorm van $P(j, b)$

Als laatste zullen we $P(j, b)$ uit (A.13) zover als mogelijk uitrekenen. We kunnen (A.13) herschrijven tot:

$$\begin{aligned}
 P(j, b) = & \lim_{n \rightarrow 0} \int_{-\infty}^{+\infty} \prod_a d\mu(J_a) d\mu(B_a) \delta(J_{a=1} - j) \delta(B_{a=1} - b) \cdot \\
 & \cdot \exp \left[-\frac{1}{2} (2\hat{f} + \hat{q}) \sum_a (J_a)^2 - \frac{\hat{Q}}{2} \sum_a (B_a)^2 + \frac{\hat{q}}{2} (\sum_a J_a)^2 + \right. \\
 & \left. + \frac{\hat{Q}}{2} (\sum_a B_a)^2 + \hat{r} (\sum_a J_a) (\sum_b B_b) \right]. \tag{A.49}
 \end{aligned}$$

Op de laatste 3 termen in de exponentiele kan men weer (E.25) toepassen en met de reeds gebruikte substitutie (A.19) vindt men in de limiet voor $n \rightarrow 0$:

$$P(j, b) = \int_{-\infty}^{+\infty} D\hat{\gamma}(\varphi, \psi) \mathcal{J}(\varphi) \mathcal{B}(\psi), \tag{A.50}$$

hierin wordt $\mathcal{J}$ gegeven door:

$$\mathcal{J}(\varphi) = \frac{p_J(j) e^{-\frac{1}{2} (2\hat{f} + \hat{q}) (j - \frac{\sqrt{\hat{q}}}{2\hat{f} + \hat{q}} \varphi)^2}}{\int_{-\infty}^{+\infty} d\mu(J) \cdot e^{-\frac{1}{2} (2\hat{f} + \hat{q}) (J - \frac{\sqrt{\hat{q}}}{2\hat{f} + \hat{q}} \varphi)^2}} \tag{A.51}$$

en $\mathcal{B}$ door:

$$\mathcal{B}(\psi) = \frac{p_B(b) e^{-\frac{\hat{Q}}{2} (b - \frac{\psi}{\sqrt{\hat{Q}}})^2}}{\int_{-\infty}^{+\infty} d\mu(B) \cdot e^{-\frac{\hat{Q}}{2} (B - \frac{\psi}{\sqrt{\hat{Q}}})^2}}. \tag{A.52}$$

Hierbij wordt met $p_J(j)$ bedoeld de distributie die het minimalisatiedomein voor het $\mathbf{J}$ -systeem bepaald, geëvalueerd in j . Analoog voor $p_B(b)$.

A.8 Wat onder globale constraints?

In deze korte paragraaf zal getoond worden hoe men eenvoudig bovenstaande zadelpuntvergelijkingen moet aanpassen indien de precursor leerregel een minimalisatiedomein heeft bepaald door een distributie $p(\mathbf{J}) = \delta(\mathbf{J}^2 - N)$ die *niet* kan geschreven worden onder de vorm $p(\mathbf{J}) = \prod_i p(J_i)$. Teruggaande naar (A.1) zien we dan dat $d\mu(\mathbf{J}^a) = \prod_j d\mu(J_j^a)$ moet vervangen worden door $(\prod_j dJ_j^a)\delta((\mathbf{J}^a)^2 - N)$. Dan blijft de hele berekening analoog tot formule (A.8). Daar is het niet nodig een Dirac-delta functie in te voeren voor de ordeparameter f_a vermits deze 1 is. We kunnen dus in al de uitdrukkingen die volgen gewoon $f_a = 1$ stellen en niet langer extremen naar f . In dezelfde uitdrukking staat er nu echter een nieuwe Dirac-delta functie, namelijk die van de distributie $p(\mathbf{J}^a) = \delta((\mathbf{J}^a)^2 - N)$. Als we deze Dirac delta schrijven met behulp van zijn integraal representatie (E.21) krijgt men:

$$\delta((\mathbf{J}^a)^2 - N) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} d\hat{f}_a \exp[i\hat{f}_a(N - (\mathbf{J}^a)^2)] \quad (\text{A.53})$$

Hieruit volgt onmiddellijk dat de uitdrukking voor G_1 en G_o onveranderd blijft behalve dat $d\mu(J_a)$ vervangen is door dJ_a . Voor de rest blijft de berekening onveranderd. Samengevat kan men dus de bekomen zadelpuntvergelijkingen en de uitdrukking voor $P(j, b)$ zeer eenvoudig aanpassen door volgende "transformatie" uit te voeren:

$$\begin{aligned} f &= 1 \\ \int d\mu(J) &= \int dJ, \end{aligned} \quad (\text{A.54})$$

en tevens alle extrematies naar f weg te laten.

A. Replica symmetrie berekening voor $P(j, b)$

Bijlage B

$V(\lambda)$ /MSB op de hypersfeer

In deze bijlage zijn de berekeningen geschetst voor een precursor op de hypersfeer. De potentiaal $V(\lambda)$ is dus nog willekeurig gehouden (behalve natuurlijk de eis dat hij een uniek minimum heeft op de hypersfeer).

B.1 De zadelpuntvergelijkingen voor $V(\lambda)$

Allereerst zal ik de zadelpuntvergelijkingen voor de precursor afleiden. We moeten dan terug naar de uitdrukkingen (2.9–2.12). Let wel, deze uitdrukkingen zijn afgeleid voor *niet* globale constraints en we moeten dus de transformatie (2.23) uitvoeren. Tevens moeten we de extrematie naar de parameter f buiten beschouwing laten. Dit laatste is echter een probleem bij deze vergelijkingen. Deze zadelpuntvergelijkingen zijn immers bekomen door extrematie naar q , $(f - q)$, $\hat{q}$ en $(2\hat{f} + \hat{q})$ (zie Bijlage A). Indien deze vergelijkingen zouden bekomen zijn door rechtstreeks af te leiden naar q , f , $\hat{q}$ en $\hat{f}$ volstond het van de vergelijking bekomen door extrematie naar f te laten vallen en enkel de andere vergelijkingen te beschouwen. In dit geval is het echter niet zo eenvoudig om te beslissen welke vergelijking weg te laten. We zullen dan ook een stap terug nemen en de te extremeren uitdrukking voor het precursor perceptron opschrijven. Dit volgt eenvoudig uit (A.24) als we er het precursor systeem uit selecteren en de transformatie (2.23) er op uitvoeren:

$$\text{Extr}_{\{q, \hat{q}, \hat{f}\}} \left\{ \hat{f} + \frac{1}{2} q \hat{q} + \right.$$

B. $V(\lambda)$ /MSB op de hypersfeer

$$\begin{aligned}
 & + \alpha \int_{-\infty}^{+\infty} D\sigma \ln \left[\int_{-\infty}^{+\infty} \frac{d\lambda}{\sqrt{2\pi(1-q)}} e^{-\beta V(\lambda) - \frac{1}{2(1-q)}(\lambda - \sqrt{q}\sigma)^2} \right] + \\
 & + \int_{-\infty}^{+\infty} D\varphi \ln \left[\int_{-\infty}^{+\infty} dJ \exp \left[-\frac{1}{2}(2\hat{f} + \hat{q})J^2 + \sqrt{\hat{q}}J\varphi \right] \right] \Bigg\}.
 \end{aligned} \tag{B.1}$$

De laatste integraal kan uitgerekend worden met de Gaussische integratie (E.10) zodat de uitdrukking wordt:

$$\begin{aligned}
 & \text{Extr}_{\{q, \hat{q}, \hat{f}\}} \left\{ \frac{1}{2}(2\hat{f} + \hat{q}) - \frac{\hat{q}}{2}(1-q) + \right. \\
 & + \alpha \int_{-\infty}^{+\infty} D\sigma \ln \left[\int_{-\infty}^{+\infty} \frac{d\lambda}{\sqrt{2\pi(1-q)}} e^{-\beta V(\lambda) - \frac{1}{2(1-q)}(\lambda - \sqrt{q}\sigma)^2} \right] + \\
 & \left. + \frac{1}{2} \ln \frac{2\pi}{2\hat{f} + \hat{q}} + \frac{\hat{q}}{2(2\hat{f} + \hat{q})} \right\}.
 \end{aligned} \tag{B.2}$$

Afleidende naar $(2\hat{f} + \hat{q})$ en $\hat{q}$ levert twee algebraïsche vergelijkingen:

$$\frac{2(\hat{f} + \hat{q})}{(2\hat{f} + \hat{q})^2} = 1; \quad (2\hat{f} + \hat{q}) = \frac{1}{1-q} \tag{B.3}$$

Hieruit volgt onmiddellijk dat $\hat{q} = \frac{1}{1-q} - 2\hat{f}$ en dat $2\hat{f} = (2\hat{f} + \hat{q})^2 - 2\hat{q} = \frac{1}{(1-q)^2} - 2\hat{q}$. Zodat er geldt:

$$\hat{q} = \frac{q}{(1-q)^2} \tag{B.4}$$

Aldus kunnen we $(2\hat{f} + \hat{q})$ en $\hat{q}$ in (B.2) substitueren zodat de vergelijking enkel nog afhangt van q . Aldus bekomen we:

$$\begin{aligned}
 & \text{Extr}_q \left\{ \frac{1}{2} \frac{1}{1-q} + \frac{1}{2} \ln 2\pi + \ln(1-q) \right. \\
 & \left. + \alpha \int_{-\infty}^{+\infty} D\sigma \ln \left[\int_{-\infty}^{+\infty} \frac{d\lambda}{\sqrt{2\pi(1-q)}} e^{-\beta \left(V(\lambda) - \frac{1}{2x}(\lambda - \sqrt{q}\sigma)^2 \right)} \right] \right\}.
 \end{aligned} \tag{B.5}$$

B.2. De zadelpuntvergelijkingen voor r en $\hat{r}$

Hierbij is $x = \beta(1 - q)$. In de limiet voor $\beta \rightarrow \infty$ zal $q \rightarrow 1$ en wordt de extrematie naar q een extrematie naar x . Als we laatste vergelijking vermenigvuldigen met $\frac{1}{\beta}$ vinden we in die limiet:

$$\text{Extr}_x \left\{ \frac{1}{2x} - \alpha \int_{-\infty}^{+\infty} D\sigma \left[V(\lambda_o(\sigma, x)) + \frac{1}{2x}(\lambda_o(\sigma, x) - \sigma)^2 \right] \right\} \quad (\text{B.6})$$

waarbij $\lambda_o(\sigma, x)$ bepaald wordt door:

$$\lambda_o(\sigma, x) = \text{Arg Min}_{\lambda} \left[V(\lambda) + \frac{(\lambda - \sigma)^2}{2x} \right] \quad (\text{B.7})$$

Extremeren naar x levert dan onmiddellijk voor de zadelpuntvergelijking voor x :

$$1 = \alpha \int_{-\infty}^{+\infty} D\sigma [\lambda_o(\sigma, x) - \sigma]^2 \quad (\text{B.8})$$

Dit is de enige zadelpuntvergelijking voor het precursor systeem.

B.2 De zadelpuntvergelijkingen voor r en $\hat{r}$

In dit geval kunnen we zonder problemen de transformatie (2.23) toepassen op de vergelijkingen (2.17) en (2.18). Zo bekomen we voor de vergelijking voor r in de limiet $\beta \rightarrow \infty$ (en dus $(2\hat{f} + \hat{q}) \rightarrow \infty$) met de methode van Laplace:

$$r = \int_{-\infty}^{+\infty} D_{\hat{\gamma}}(\varphi, \psi) \frac{\sqrt{\hat{q}}}{2\hat{f} + \hat{q}} \varphi \tanh(\sqrt{\hat{Q}}\psi) \quad (\text{B.9})$$

Met achtereenvolgens gebruik te maken van de eigenschappen (E.20) en (E.15) van de Gaussische integratiematen hebben we dan:

$$r = \hat{\gamma} \int_{-\infty}^{+\infty} D\psi \psi \tanh(\sqrt{\hat{Q}}\psi) = \hat{\gamma} \int_{-\infty}^{+\infty} \frac{\sqrt{\hat{Q}}}{\cosh^2(\sqrt{\hat{Q}}\psi)} \quad (\text{B.10})$$

Gebruik makend van de zadelpuntvergelijking voor Q (2.14) levert dit:

$$r = \hat{\gamma} \sqrt{\hat{Q}}(1 - Q) = \frac{\hat{r}}{\sqrt{\hat{q}}}(1 - Q) \quad (\text{B.11})$$

B. $V(\lambda)$ /MSB op de hypersfeer

Tenslotte is er nog de zadelpuntvergelijking voor $\hat{r}$ (2.18). Deze kan weerom onmiddellijk met de methode van Laplace voor $\beta \rightarrow \infty$ geschreven worden als:

$$\hat{r} = \alpha \int_{-\infty}^{+\infty} D_\gamma(\tau, \sigma) \left[\frac{\lambda_o(\sigma, x) - \sigma}{1 - q} \right] \cdot \frac{\frac{1}{\sqrt{2\pi(1-Q)}} \cdot e^{-\frac{(\kappa_b - \sqrt{Q}\tau)^2}{2(1-Q)}}}{H\left(\frac{\kappa_b - \sqrt{Q}\tau}{\sqrt{1-Q}}\right)} \quad (\text{B.12})$$

Vermits de vergelijking (B.11) algebraïsch is kan men in (B.12) $\hat{r}$ vervangen door $\frac{r\sqrt{\hat{q}}}{1-Q}$. Zo bekomt men dan 1 vergelijking voor r :

$$\begin{aligned} \sqrt{1-Q} \frac{r\sqrt{\hat{q}}}{1-Q} = & \alpha \int_{-\infty}^{+\infty} \frac{d\sigma}{2\pi} e^{-\frac{1}{2(1-\gamma^2)}\sigma^2} \cdot \left[\frac{\lambda_o(\sigma, x) - \sigma}{1 - q} \right] \cdot \\ & \int_{-\infty}^{+\infty} \frac{d\tau}{\sqrt{2\pi(1-\gamma^2)}} \cdot e^{-\frac{1}{2(1-\gamma^2)}[\tau^2 - 2\gamma\sigma\tau]} \cdot \frac{e^{-\frac{(\kappa_b - \sqrt{Q}\tau)^2}{2(1-Q)}}}{H\left(\frac{\kappa_b - \sqrt{Q}\tau}{\sqrt{1-Q}}\right)} \end{aligned} \quad (\text{B.13})$$

Als we de van de term $[\tau^2 - 2\gamma\sigma\tau]$ in de exponent van de laatste integraal een volledig kwadraat maken volgt er:

$$\begin{aligned} \frac{r\sqrt{\hat{q}}}{\sqrt{1-Q}} = & \alpha \int_{-\infty}^{+\infty} \frac{d\sigma}{2\pi} e^{-\frac{1}{2}\sigma^2} \cdot \left[\frac{\lambda_o(\sigma, x) - \sigma}{1 - q} \right] \cdot \\ & \int_{-\infty}^{+\infty} \frac{d\tau}{\sqrt{2\pi(1-\gamma^2)}} \cdot e^{-\frac{1}{2(1-\gamma^2)}[\tau - \gamma\sigma]^2} \cdot \frac{e^{-\frac{(\kappa_b - \sqrt{Q}\tau)^2}{2(1-Q)}}}{H\left(\frac{\kappa_b - \sqrt{Q}\tau}{\sqrt{1-Q}}\right)} \end{aligned} \quad (\text{B.14})$$

Als we nu noteren:

$$g_r(\sigma) = \sqrt{1-Q} \int_{-\infty}^{+\infty} \frac{d\tau}{\sqrt{2\pi(1-\gamma^2)}} \cdot e^{-\frac{1}{2(1-\gamma^2)}[\tau - \gamma\sigma]^2} \cdot \frac{1}{\sqrt{2\pi}} \frac{e^{-\frac{(\kappa_b - \sqrt{Q}\tau)^2}{2(1-Q)}}}{H\left(\frac{\kappa_b - \sqrt{Q}\tau}{\sqrt{1-Q}}\right)} \quad (\text{B.15})$$

en opmerken dat, wegens (B.4) $\sqrt{\hat{q}}(1-q) \rightarrow 1$ voor $\beta \rightarrow \infty$ dan kunnen volgt er voor de zadelpuntvergelijking voor r :

$$r = \alpha \int_{-\infty}^{+\infty} D\sigma [\lambda_o(\sigma, x) - \sigma] \cdot g_r(\sigma) \quad (\text{B.16})$$

B.3 Berekening van $P(j, b)$ en de kwaliteitsmaat f_x

De distributie $P(j, b)$ volgt vrijwel onmiddellijk uit de algemene uitdrukking (2.20...2.22). Rekening houdend met de voorwaarde voor globale constraints (2.23) en dat $(2\hat{f} + \hat{q}) \rightarrow \infty$ voor $\beta \rightarrow \infty$ vindt men voor $\mathcal{J}(\varphi)$ met behulp van de representatie van de Dirac-delta functie (E.23):

$$\mathcal{J}(\varphi) = \delta(j - \varphi) \quad (\text{B.17})$$

en voor $\mathcal{W}$ geldt er:

$$\mathcal{W}(\psi) = \frac{1}{2} [\delta(b+1) + \delta(b-1)] \cdot \frac{e^{-\frac{\hat{Q}}{2}(b - \frac{\psi}{\sqrt{\hat{Q}}})^2}}{\frac{1}{2} e^{-\frac{1}{2}\hat{Q}} \left[e^{(1 - \frac{\psi}{\sqrt{\hat{Q}}})^2} + e^{(1 + \frac{\psi}{\sqrt{\hat{Q}}})^2} \right]}. \quad (\text{B.18})$$

Uitwerken van de kwadraten in de exponentiëlen en vereenvoudigen van teller en noemer levert:

$$P(j, b) = \int_{-\infty}^{+\infty} D\gamma(\varphi, \psi) \delta(j - \varphi) \frac{1}{2} [\delta(b+1) + \delta(b-1)] \cdot \frac{e^{\sqrt{\hat{Q}} b \psi}}{\cosh(\sqrt{\hat{Q}} \psi)} \quad (\text{B.19})$$

Met deze uitdrukking voor $P(j, b)$ kan men nu de fractie van componenten $|J| > J_o > 0$ berekenen die bij knippen correct de corresponderende component van de MSB voorspellen. Vermits $P(j, b) = P(-j, -b)$ kunnen we deze fractie schrijven als:

$$f_{J_o} = \frac{\int_{J_o}^{+\infty} dj P(j, +1)}{\int_{J_o}^{+\infty} dj P(j)}, \quad (\text{B.20})$$

waarbij $P(j) = \int_{-\infty}^{+\infty} db P(j, b)$. Bemerk dat als $J_o = 0$ de grootte $f_{J_o=0}$ gelijk is aan de fractie van componenten die correct de overeenkomstige component van de MSB voorspellen bij volledig knippen van de precursor.

Laten we eerst de teller van (B.20) berekenen. Deze is gelijk aan:

B. $V(\lambda)$ /MSB op de hypersfeer

$$\frac{1}{2} \int_{-\infty}^{+\infty} D\hat{\gamma}(\varphi, \psi) \frac{e^{\sqrt{\hat{Q}}\psi}}{\cosh(\sqrt{\hat{Q}}\psi)} \int_{J_o}^{+\infty} dj \delta(j - \varphi) \quad (\text{B.21})$$

De integraal over j is gelijk aan $\Theta[\varphi - J_o]$ zodat er volgt:

$$\frac{1}{2} \int_{-\infty}^{+\infty} D\psi \int_{J_o}^{+\infty} \frac{d\varphi}{\sqrt{2\pi(1 - \hat{\gamma}^2)}} e^{-\frac{1}{2} \frac{(\varphi - \hat{\gamma}\psi)^2}{1 - \hat{\gamma}^2}} \cdot \frac{e^{\sqrt{\hat{Q}}\psi}}{\cosh(\sqrt{\hat{Q}}\psi)} \quad (\text{B.22})$$

Of nog:

$$\frac{1}{2} \int_{-\infty}^{+\infty} D\psi \frac{e^{\sqrt{\hat{Q}}\psi}}{\cosh(\sqrt{\hat{Q}}\psi)} H\left(\frac{J_o - \hat{\gamma}\psi}{\sqrt{1 - \hat{\gamma}^2}}\right) \quad (\text{B.23})$$

Men kan dit nog herschrijven als:

$$\frac{1}{4} \int_{-\infty}^{+\infty} D\psi \frac{e^{\sqrt{\hat{Q}}\psi} + e^{-\sqrt{\hat{Q}}\psi}}{\cosh(\sqrt{\hat{Q}}\psi)} H\left(\frac{J_o - \hat{\gamma}\psi}{\sqrt{1 - \hat{\gamma}^2}}\right). \quad (\text{B.24})$$

Indien we in de teller van het integrandum een term $e^{-\sqrt{\hat{Q}}\psi}$ toevoegen en ook weer aftrekken dan vinden we:

$$\frac{1}{2} \int_{-\infty}^{+\infty} D\psi \left[1 + \tanh(\sqrt{\hat{Q}}\psi)\right] H\left(\frac{J_o - \hat{\gamma}\psi}{\sqrt{1 - \hat{\gamma}^2}}\right). \quad (\text{B.25})$$

Rest er ons enkel nog de noemer van (B.20) te bepalen:

$$\int_{J_o}^{+\infty} P(j) dj = \int_{J_o}^{+\infty} \int_{-\infty}^{+\infty} P(j, b) db \quad (\text{B.26})$$

Met behulp van de uitdrukking (B.19) wordt dit:

$$\begin{aligned} \int_{J_o}^{+\infty} P(j) dj &= \int_{J_o}^{+\infty} \int_{-\infty}^{+\infty} D\hat{\gamma}(\varphi, \psi) \delta(j - \varphi) \\ &= \int_{J_o}^{+\infty} \int_{-\infty}^{+\infty} D\varphi \delta(j - \varphi) \\ &= H(J_o) \end{aligned} \quad (\text{B.27})$$

B.3. Berekening van $P(j, b)$ en de kwaliteitsmaat f_x

Hierbij is tevens gebruik gemaakt van de eigenschap (E.19) van de twee dimensionale integratiemaat.

Merk op dat invullen van (B.19) in (B.26) twee integralen levert waarvan er een identiek is aan de teller (B.23) en de andere ervan verschilt met een teken in de exponent. De noemer kan dus ook geschreven worden als:

$$\int_{-\infty}^{+\infty} D\psi H\left(\frac{J_o - \hat{\gamma}\psi}{\sqrt{1 - \hat{\gamma}^2}}\right) = H(J_o). \quad (\text{B.28})$$

Waarbij de laatste gelijkheid volgt uit de uitdrukking voor de noemer in (B.27). Met behulp van (B.28) kunnen we nu (B.20) schrijven als:

$$f_{J_o} = \frac{1}{2} + \frac{1}{2H(J_o)} \int_{-\infty}^{+\infty} D\psi \tanh(\sqrt{\hat{Q}}\psi) H\left(\frac{J_o - \hat{\gamma}\psi}{\sqrt{1 - \hat{\gamma}^2}}\right). \quad (\text{B.29})$$

Om de kwaliteitsmaat f_x van (2.2) te bekomen, die de fractie van correct voorspelde componenten bepaalt als er xN componenten geknipt zijn, volstaat het J_o te bepalen uit $x = 2 \int_{J_o}^{+\infty} P(j) dj = 2 H(J_o)$.

B. $V(\lambda)/\text{MSB}$ op de hypersfeer

Bijlage C

$V(\lambda)$ /MSB in de hyperkubus

In deze bijlage zijn de berekeningen geschetst voor een precursor in de hyperkubus. De potentiaal $V(\lambda)$ is dus nog willekeurig gehouden. (Behalve natuurlijk de eis dat hij een uniek minimum heeft in de hyperkubus.)

C.1 De zadelpuntvergelijkingen voor $V(\lambda)$

Allereerst zal ik de zadelpuntvergelijkingen voor de precursor in de hyperkubus afleiden. We moeten dan terug naar de uitdrukkingen (2.9–2.12). De distributie die het minimalisatiedomein definieert is gegeven door $p(J_i) = \frac{1}{2}\Theta(1 - J_i^2)$ zodat er geldt voor de uitdrukkingen (2.11) en (2.12):

$$-\frac{1}{2}(f - q) + \int_{-\infty}^{+\infty} D\varphi \frac{\int_{-1}^{+1} dJ e^{-(2\hat{f}+\hat{q})\left[\frac{1}{2}J^2 - \frac{\sqrt{\hat{q}}}{2\hat{f}+\hat{q}}J\varphi\right]} \cdot \frac{J\varphi}{2\sqrt{\hat{q}}}}{\int_{-1}^{+1} dJ e^{-(2\hat{f}+\hat{q})\left[\frac{1}{2}J^2 - \frac{\sqrt{\hat{q}}}{2\hat{f}+\hat{q}}J\varphi\right]}} = 0 \quad (\text{C.1})$$

$$\frac{1}{2}f + \int_{-\infty}^{+\infty} D\varphi \frac{\int_{-1}^{+1} dJ e^{-(2\hat{f}+\hat{q})\left[\frac{1}{2}J^2 - \frac{\sqrt{\hat{q}}}{2\hat{f}+\hat{q}}J\varphi\right]} \cdot \left(-\frac{1}{2}J^2\right)}{\int_{-1}^{+1} dJ e^{-(2\hat{f}+\hat{q})\left[\frac{1}{2}J^2 - \frac{\sqrt{\hat{q}}}{2\hat{f}+\hat{q}}J\varphi\right]}} = 0. \quad (\text{C.2})$$

In de limiet voor $\beta \rightarrow \infty$ gaat $q \rightarrow f$ en $(2\hat{f} + \hat{q}) \rightarrow \infty$. Met de substitutie $\frac{\sqrt{\hat{q}}}{2\hat{f}+\hat{q}} = t$ en t eindig, wordt dit dan met behulp van de methode van Laplace:

C. $V(\lambda)/\text{MSB}$ in de hyperkubus

$$(f - q) = \frac{1}{\sqrt{\hat{q}}} \int_{-\infty}^{+\infty} D\varphi \varphi J_{\min}(t\varphi), \quad (\text{C.3})$$

$$f = \int_{-\infty}^{+\infty} D\varphi [J_{\min}(t\varphi)]^2. \quad (\text{C.4})$$

Waarbij $J_{\min}(t\varphi)$ gegeven is door:

$$J_{\min}(t\varphi) = \text{Arg Min}_{J \in [-1,1]} \left[\frac{1}{2} J^2 - t J \varphi \right] = \begin{cases} t\varphi & : \text{als } |t\varphi| \leq 1 \\ +1 & : \text{als } t\varphi > 1 \\ -1 & : \text{als } t\varphi < -1 \end{cases}. \quad (\text{C.5})$$

De zadelpuntvergelijkingen (2.9) en (2.10) blijven onveranderd:

$$(2\hat{f} + \hat{q}) + \frac{\alpha}{(f - q)\sqrt{f}} \int_{-\infty}^{+\infty} D\sigma \sigma [\lambda_o(\sqrt{f}\sigma, x) - \sqrt{f}\sigma] = 0, \quad (\text{C.6})$$

$$\hat{f} - \frac{\alpha}{2(f - q)} + \frac{\alpha}{2(f - q)^2} \int_{-\infty}^{+\infty} D\sigma [\lambda_o(\sqrt{f}\sigma, x) - \sqrt{f}\sigma]^2 = 0. \quad (\text{C.7})$$

Hierin is λ_o gegeven door:

$$\lambda_o(\sqrt{f}\sigma, x) = \text{Arg Min}_{\lambda} \left[V(\lambda) + \frac{(\lambda - \sqrt{f}\sigma)^2}{2x} \right], \quad (\text{C.8})$$

met $x = \beta(f - q)$. Als we verder nog de substitutie $y = (f - q)(2\hat{f} + \hat{q})$ invoeren kunnen we de oneindigheden in de zadelpuntvergelijkingen weg werken en bekomen we voor (C.6) na vermenigvuldigen met $(f - q)$:

$$y = -\alpha \int_{-\infty}^{+\infty} D\sigma \left[\frac{\partial \lambda_o}{\partial \sqrt{f}\sigma} (\sqrt{f}\sigma, x) - 1 \right], \quad (\text{C.9})$$

waar we tevens gebruik hebben gemaakt van (E.15). Voor (C.7) vinden we na vermenigvuldiging met $(f - q)^2$:

$$(ty)^2 = \alpha \int_{-\infty}^{+\infty} D\sigma [\lambda_o(\sqrt{f}\sigma, x) - \sqrt{f}\sigma]^2, \quad (\text{C.10})$$

C.2. De zadelpuntvergelijkingen voor r en $\hat{r}$

vermits $-2\hat{f}(f-q)^2 = -(2\hat{f}+\hat{q})(f-q)^2 + \hat{q}(f-q)^2 = -y(f-q) + (yt)^2 = (yt)^2$. Verder geldt er voor (C.3):

$$yt = \int_{-\infty}^{+\infty} D\varphi \frac{\partial J_{\min}}{\partial \varphi}(t\varphi) = t \int_{-1/t}^{+1/t} D\varphi \quad (\text{C.11})$$

of nog

$$y = H(-1/t) - H(1/t) = \text{Erf}\left(\frac{1}{t\sqrt{2}}\right), \quad (\text{C.12})$$

waarbij gebruik gemaakt is van (E.7) en het feit dat de error functie een oneven functie is. Als laatste volgt er voor de zadelpuntvergelijking (C.4):

$$\begin{aligned} f &= \int_{-\infty}^{-1/t} D\varphi + \int_{-1/t}^{+1/t} D\varphi (t\varphi)^2 + \int_{1/t}^{+\infty} D\varphi \\ &= 2 \int_{1/t}^{+\infty} D\varphi + t^2 \left[\int_{-1/t}^{+\infty} D\varphi \varphi^2 - \int_{1/t}^{+\infty} D\varphi \varphi^2 \right]. \end{aligned} \quad (\text{C.13})$$

*Als we gebruik maken van (E.17) vinden we:

$$\begin{aligned} f &= 2H(1/t) + t^2 \left[\frac{-2}{\sqrt{2\pi}} \frac{1}{t} \exp\left[\frac{-1}{2t^2}\right] + H(-1/t) - H(1/t) \right] \\ &= 2H(1/t) - \frac{2t}{\sqrt{2\pi}} \exp\left[\frac{-1}{2t^2}\right] + yt^2. \end{aligned} \quad (\text{C.14})$$

Gebruik makend van de definitie van de H-functie (E.7) en (C.12) vinden we dan:

$$f = 1 - y + yt^2 - \sqrt{\frac{2}{\pi}} t \exp\left[\frac{-1}{2t^2}\right] \quad (\text{C.15})$$

C.2 De zadelpuntvergelijkingen voor r en $\hat{r}$

We gaan terug naar de vergelijkingen (2.17) en (2.18) waar we de distributie $p(J_i) = \frac{1}{2}\Theta(1 - J_i^2)$ invullen. Zo vindt men voor (2.17) voor $\beta \rightarrow \infty$, $q \rightarrow f$ en $2\hat{f} + \hat{q} \rightarrow \infty$ met de methode van Laplace:

C. $V(\lambda)$ /MSB in de hyperkubus

$$r = \int_{-\infty}^{+\infty} D_{\hat{\gamma}}(\varphi, \psi) \cdot \tanh(\sqrt{\hat{Q}}\psi) \cdot J_{\min}(t\varphi), \quad (\text{C.16})$$

hierbij is, $J_{\min}$ gegeven door (C.5) en $t = \frac{\sqrt{\hat{q}}}{2\hat{f}+\hat{q}}$ zoals in vorige sectie. Voor vergelijking (2.18) volgt in de limiet voor $\beta \rightarrow \infty$ op analoge wijze als reeds berekend in sectie B.2:

$$\hat{r}(1-Q)(f-q) = \alpha \int_{-\infty}^{+\infty} D\sigma \left[\lambda_o(\sqrt{f}\sigma, x) - \sqrt{f}\sigma \right] \cdot g_r(\sigma). \quad (\text{C.17})$$

Hierbij is $g_r(\sigma)$ is bepaald door (B.15). Vermits $\hat{\gamma} = \frac{\hat{r}}{\sqrt{\hat{q}\hat{Q}}}$ kunnen we dit nog herschrijven tot:

$$\hat{\gamma}\sqrt{\hat{Q}}(1-Q)ty = \alpha \int_{-\infty}^{+\infty} D\sigma \left[\lambda_o(\sqrt{f}\sigma, x) - \sqrt{f}\sigma \right] \cdot g_r(\sigma). \quad (\text{C.18})$$

C.2.1 Betekenis van de zadelpuntvergelijking voor r

De zadelpuntvergelijking (4.9) drukt uit dat r de gemiddelde waarde is van het product van de overeenkomstige componenten¹ van $\mathbf{J}$ en $\mathbf{B}$: $\langle bj \rangle = \sum_b \int dj P(j, b) bj$. Dit kan eenvoudig aangetoond worden m. b. v de uitdrukking voor $P(j, b)$ bekomen in sectie C.3:

$$P(j, b) = \int_{-\infty}^{+\infty} D_{\hat{\gamma}}(\varphi, \psi) \cdot \Delta(j, t\varphi) \frac{1}{2} \left[1 + \tanh(\sqrt{\hat{Q}}b\psi) \right], \quad (\text{C.19})$$

waarbij:

$$\Delta(j, t\varphi) = \begin{cases} \delta(j - t\varphi) & : \text{als } |t\varphi| \leq 1 \\ \delta(j - 1) & : \text{als } t\varphi > 1 \\ \delta(j + 1) & : \text{als } t\varphi < -1. \end{cases} \quad (\text{C.20})$$

Nu volgt er vrijwel onmiddellijk dat $\langle bj \rangle$ gegeven wordt door:

$$\int_{-\infty}^{+\infty} D_{\hat{\gamma}}(\varphi, \psi) \cdot J_{\min}(t\varphi) \cdot \sum_b b \frac{1}{2} \left[1 + \tanh(\sqrt{\hat{Q}}b\psi) \right], \quad (\text{C.21})$$

vermits $J_{\min}(t\varphi) = \int_{-1}^{+1} dj j \Delta(j, t\varphi)$, waarbij $J_{\min}(t\varphi)$ gegeven is door (C.5). Uitwerken van de sommatie over b levert dan dat $\langle bj \rangle = r$.

¹Merk op dat dit tevens geldig is voor de zadelpuntvergelijking (B.9) voor r op de hypersfeer. Een gelijkaardige redenering als dewelke hier volgt maar waar men gebruik maakt van de uitdrukking (B.19) voor $P(j, b)$ op de hypersfeer levert dit resultaat.

C.3 Berekening van $P(j, b)$ en de kwaliteitsmaat f_x

De distributie $P(j, b)$ volgt uit de algemene uitdrukking (2.20...2.22). Rekening houdend met $(2f + \hat{q}) \rightarrow \infty$ voor $\beta \rightarrow \infty$ vindt men voor $\mathcal{J}(\varphi)$ met behulp van de representatie van de Dirac-delta functie (E.24):

$$\mathcal{J}(\varphi) = \Theta(1 - j^2) \Delta(j, t\varphi), \quad (\text{C.22})$$

waarbij $\Delta(j, t\varphi)$ gegeven is door:

$$\Delta(j, t\varphi) = \begin{cases} \delta(j - t\varphi) & : \text{als } |t\varphi| \leq 1 \\ \delta(j - 1) & : \text{als } t\varphi > 1 \\ \delta(j + 1) & : \text{als } t\varphi < -1 \end{cases} \quad (\text{C.23})$$

en voor $\mathcal{W}$ geldt er analoog als in sectie B.3:

$$\mathcal{W}(\psi) = \frac{1}{2} [\delta(b + 1) + \delta(b - 1)] \cdot \frac{e^{\sqrt{\hat{Q}}b\psi}}{\cosh(\sqrt{\hat{Q}}\psi)} \quad (\text{C.24})$$

Om de berekening van $P(j, b)$ uit te voeren merken we verder op dat:

$$P(j, b) = \frac{1}{2} \delta(b + 1) P(j| - 1) + \frac{1}{2} \delta(b - 1) P(j| + 1), \quad (\text{C.25})$$

waarbij $P(j|\pm 1)$ de conditionele waarschijnlijkheden zijn. Met behulp van (C.22) en (C.24) kunnen we deze schrijven als:

$$\begin{aligned} P(j|\pm 1) = \Theta(1 - j^2) & \left[\delta(j + 1) \int_{-\infty}^{-1/t} \int_{-\infty}^{+\infty} D_{\hat{\gamma}}(\varphi, \psi) \frac{e^{\pm\sqrt{\hat{Q}}\psi}}{\cosh(\sqrt{\hat{Q}}\psi)} + \right. \\ & + \delta(j - 1) \int_{1/t}^{+\infty} \int_{-\infty}^{+\infty} D_{\hat{\gamma}}(\varphi, \psi) \frac{e^{\pm\sqrt{\hat{Q}}\psi}}{\cosh(\sqrt{\hat{Q}}\psi)} + \\ & \left. + \int_{-1/t}^{+1/t} \int_{-\infty}^{+\infty} D_{\hat{\gamma}}(\varphi, \psi) \frac{1}{t} \delta(\varphi - j/t) \cdot \frac{e^{\pm\sqrt{\hat{Q}}\psi}}{\cosh(\sqrt{\hat{Q}}\psi)} \right] \quad (\text{C.26}) \end{aligned}$$

C. $V(\lambda)$ /MSB in de hyperkubus

Hierbij zijn de integraties die niet lopen van $-\infty$ tot $+\infty$ de integraties over de variabele φ . Met de uitdrukking (C.25) voor $P(j, b)$ kan men nu de fractie van componenten $|J| > J_o > 0$ berekenen die bij knippen correct de corresponderende component van de MSB voorspellen. Vermits $P(j, b) = P(-j, -b)$ kunnen we deze fractie schrijven als:

$$f_{J_o} = \frac{\int_{J_o}^{+\infty} dj P(j, +1)}{\int_{J_o}^{+\infty} dj P(j)} = \frac{\int_{J_o}^{+\infty} dj P(j|+1)}{\int_{J_o}^{+\infty} dj P(j|+1) + \int_{J_o}^{+\infty} dj P(j|-1)} \quad (C.27)$$

Bemerk dat als $J_o = 0$ de grootheid $f_{J_o=0}$ gelijk is aan de fractie van componenten die correct de overeenkomstige component van de MSB voorspellen bij volledig knippen van de precursor. Om zowel de teller als de noemer te bepalen volstaat het van $\int_{J_o}^{+\infty} dj P(j|\pm 1)$ te berekenen. Gebruik maken van (C.26) levert dan:

$$\begin{aligned} \int_{J_o}^{+\infty} dj P(j|\pm 1) &= \int_{1/t}^{+\infty} \int_{-\infty}^{+\infty} D_{\hat{\gamma}}(\varphi, \psi) \frac{e^{\pm\sqrt{\hat{Q}}\psi}}{\cosh(\sqrt{\hat{Q}}\psi)} + \\ &+ \int_{J_o}^{+1} dj \int_{-1/t}^{+1/t} D_{\hat{\gamma}}(\varphi, \psi) \frac{1}{t} \delta(\varphi - j/t) \cdot \frac{e^{\pm\sqrt{\hat{Q}}\psi}}{\cosh(\sqrt{\hat{Q}}\psi)}. \end{aligned} \quad (C.28)$$

of nog:

$$\begin{aligned} \int_{J_o}^{+\infty} dj P(j|\pm 1) &= \int_{-\infty}^{+\infty} D\psi \frac{e^{\pm\sqrt{\hat{Q}}\psi}}{\cosh(\sqrt{\hat{Q}}\psi)} \int_{1/t}^{+\infty} \frac{d\varphi}{\sqrt{2\pi(1-\hat{\gamma}^2)}} \cdot e^{-\frac{1}{2(1-\hat{\gamma}^2)}(\varphi-\hat{\gamma}\psi)^2} + \\ &+ \int_{J_o}^{+1} dj \int_{-1/t}^{+1/t} D\varphi \frac{1}{t} \delta(\varphi - j/t) \cdot \\ &\cdot \int_{-\infty}^{+\infty} \frac{d\psi}{\sqrt{2\pi(1-\hat{\gamma}^2)}} \cdot e^{-\frac{1}{2(1-\hat{\gamma}^2)}(\psi-\hat{\gamma}\varphi)^2} \cdot \frac{e^{\pm\sqrt{\hat{Q}}\psi}}{\cosh(\sqrt{\hat{Q}}\psi)}. \end{aligned} \quad (C.29)$$

Dit kan dan weer herschreven worden tot:

C.3. Berekening van $P(j, b)$ en de kwaliteitsmaat f_x

$$\begin{aligned} \int_{J_o}^{+\infty} dj P(j|\pm 1) &= \int_{-\infty}^{+\infty} D\psi \frac{e^{\pm\sqrt{\hat{Q}}\psi}}{\cosh(\sqrt{\hat{Q}}\psi)} \cdot H\left(\frac{1/t - \hat{\gamma}\psi}{\sqrt{1 - \hat{\gamma}^2}}\right) + \\ &+ \int_{J_o/t}^{1/t} Dj \int_{-\infty}^{+\infty} Dz \frac{e^{\pm\sqrt{\hat{Q}}(z\sqrt{1 - \hat{\gamma}^2} + \hat{\gamma}j)}}{\cosh(\sqrt{\hat{Q}}(z\sqrt{1 - \hat{\gamma}^2} + \hat{\gamma}j))} \end{aligned} \quad (C.30)$$

of nog:

$$\begin{aligned} \int_{J_o}^{+\infty} dj P(j|\pm 1) &= \int_{-\infty}^{+\infty} D\psi \left[1 + \tanh(\pm\sqrt{\hat{Q}}\psi)\right] \cdot H\left(\frac{1/t - \hat{\gamma}\psi}{\sqrt{1 - \hat{\gamma}^2}}\right) + \\ &+ \int_{J_o/t}^{1/t} Dj \int_{-\infty}^{+\infty} Dz \left[1 + \tanh(\pm\sqrt{\hat{Q}}(z\sqrt{1 - \hat{\gamma}^2} + \hat{\gamma}j))\right] \end{aligned} \quad (C.31)$$

Met behulp van (B.28) vinden we dan:

$$\begin{aligned} H(J_o/t) &+ \int_{-\infty}^{+\infty} D\psi \tanh(\pm\sqrt{\hat{Q}}\psi) \cdot H\left(\frac{1/t - \hat{\gamma}\psi}{\sqrt{1 - \hat{\gamma}^2}}\right) + \\ &+ \int_{J_o/t}^{1/t} Dj \int_{-\infty}^{+\infty} Dz \tanh(\pm\sqrt{\hat{Q}}(z\sqrt{1 - \hat{\gamma}^2} + \hat{\gamma}j)) \end{aligned} \quad (C.32)$$

Laten we even enkel de laatste integraal in deze uitdrukking bekijken. Beschouw de volgende orthogonale transformatie voor deze integraal:

$$\begin{aligned} x &= z\sqrt{1 - \hat{\gamma}^2} + \hat{\gamma}j \\ y &= j\sqrt{1 - \hat{\gamma}^2} - \hat{\gamma}z \end{aligned} \quad (C.33)$$

Hiermee verandert het integratiedomein zoals geschetst in figuur (C.1).

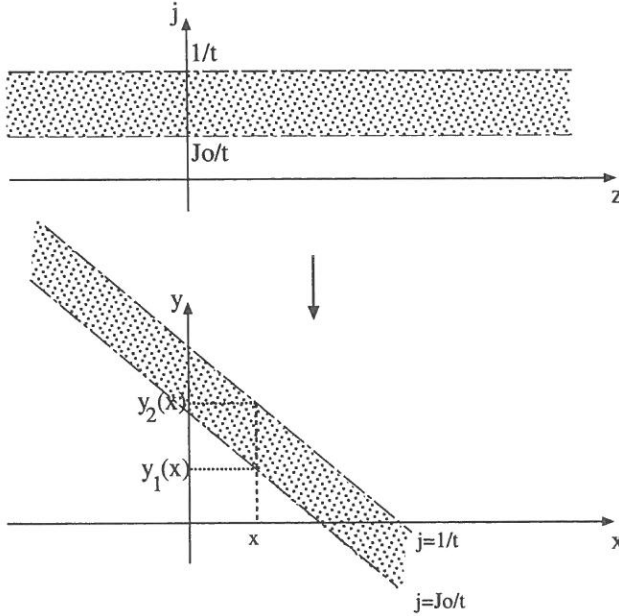
Veronderstel dat we $j = J_o/t$ vast nemen en z variëren, dan volgen we de rechte in het (z, j) -vlak bepaald door $(J_o/t, z)$. Volgens (C.33) wordt deze rechte in het (x, y) vlak gegeven door:

$$(x, y) = (z(\sqrt{1 - \hat{\gamma}^2}, -\hat{\gamma}) + (\hat{\gamma}J_o/t, \sqrt{1 - \hat{\gamma}^2}J_o/t). \quad (C.34)$$

Analoog voor een willekeurige vaste waarde van j met $J_o/t \leq j \leq 1/t$:

$$(x, y) = (z(\sqrt{1 - \hat{\gamma}^2}, -\hat{\gamma}) + (\hat{\gamma}j, \sqrt{1 - \hat{\gamma}^2}j). \quad (C.35)$$

C. $V(\lambda)$ /MSB in de hyperkubus



Figuur C.1: Verandering van het integratiedomein onder de orthogonale transformatie (C.33).

We kunnen dit nieuwe domein integreren door x te laten variëren van $-\infty$ tot $+\infty$ en voor iedere waarde van x de y variabele te integreren tussen $y_1(x)$ en $y_2(x)$, waarbij:

$$\begin{aligned} y_1(x) &= \frac{J_o/t - \hat{\gamma}x}{\sqrt{1 - \hat{\gamma}^2}} \\ y_2(x) &= \frac{1/t - \hat{\gamma}x}{\sqrt{1 - \hat{\gamma}^2}} \end{aligned} \quad (\text{C.36})$$

Zo kunnen we dan (C.32) als volgt schrijven:

$$\begin{aligned} \int_{J_o}^{+\infty} dj P(j|\pm 1) &= H(J_o/t) + \int_{-\infty}^{+\infty} Dx \tanh(\pm \sqrt{\hat{Q}}x) \int_{y_1(x)}^{y_2(x)} Dy \\ &\quad \cdot H\left(\frac{1/t - \hat{\gamma}\psi}{\sqrt{1 - \hat{\gamma}^2}}\right) + \end{aligned} \quad (\text{C.37})$$

Zodat we tenslotte kunnen schrijven voor $\int_{J_o}^{+\infty} dj P(j|\pm 1)$:

C.4. Berekening van $P(j)$ in de hyperkubus

$$H(J_o/t) + \int_{-\infty}^{+\infty} Dx \tanh(\pm\sqrt{\hat{Q}}x) H\left(\frac{J_o/t - \hat{\gamma}x}{\sqrt{1 - \hat{\gamma}^2}}\right) \quad (C.38)$$

Vermits $\tanh(x)$ een oneven functie is geldt er onmiddellijk voor de noemer van (C.27) dat deze gelijk is aan $2H(J_o/t)$ zodat :

$$f_{J_o} = \frac{1}{2} + \frac{1}{2H(J_o/t)} \int_{-\infty}^{+\infty} Dx \tanh(\sqrt{\hat{Q}}x) H\left(\frac{J_o/t - \hat{\gamma}x}{\sqrt{1 - \hat{\gamma}^2}}\right) \quad (C.39)$$

C.4 Berekening van $P(j)$ in de hyperkubus

Om de distributie $P(j)$ van de precursor componenten te bepalen moeten we terug naar uitdrukking (C.25). Hieruit bekomen we direkt:

$$P(j) = \int_{-\infty}^{+\infty} db P(j, b) = \frac{1}{2}P(j-1) + \frac{1}{2}P(j+1). \quad (C.40)$$

Met behulp van (C.26) volgt er na samenvoegen van de integraties over hetzelfde integratiedomein:

$$P(j) = \Theta(1-j^2) \frac{1}{\sqrt{2\pi t}} e^{\frac{-j^2}{2t^2}} + H(1/t)(\delta(j-1) + \delta(j+1)). \quad (C.41)$$

C.5 De kostfunctie $E(\mathbf{J})$ met de perturbatieterm $\rho \mathbf{J}^2$

In deze sectie zal ik de zadelpuntvergelijkingen afleiden van het precursor systeem als de kostfunctie $E(\mathbf{J})$ een perturbatieterm $\rho \mathbf{J}^2$ krijgt:

$$E(\mathbf{J}) = \sum_{\mu} V(\lambda_{\mu}) + \rho \mathbf{J}^2 \quad (C.42)$$

Ik zal hiervoor steunen op de reeds uitgevoerde berekeningen in bijlage A en C. In de uitdrukking (A.1) voor $P(j, b)$ treedt de precursor kostfunctie op in een exponent, zodat we nu dus een extra term $\exp[-\beta \rho \mathbf{J}^2]$ hebben. We kunnen deze term eenvoudig in de berekeningen van bijlage A opnemen door het te includeren in de integratiemaat $d\mu(\mathbf{J}^a)$:

C. $V(\lambda)/\text{MSB}$ in de hyperkubus

$$d\mu(\mathbf{J}^a) = d\mathbf{J}^a p(\mathbf{J}^a) \exp[-\beta\rho(\mathbf{J}^a)^2] \quad (\text{C.43})$$

Verder blijft dan de hele berekening ongewijzigd. Ook G_o blijft identiek aan (A.6) maar G_1 uit (A.10) wijzigt vermits $d\mu(\mathbf{J}^a)$ een extra term heeft gekregen. Rekening houdend met dit gegeven kan men de te extremeren vergelijking opstellen voor dit precursor systeem:

$$\begin{aligned} \text{Extr}_{\{q, f, \hat{q}, \hat{f}\}} \left\{ f\hat{f} + \frac{1}{2}q\hat{q} + \right. \\ \left. + \alpha \int_{-\infty}^{+\infty} D\sigma \ln \left[\int_{-\infty}^{+\infty} \frac{d\lambda}{\sqrt{2\pi(f-q)}} e^{-\beta V(\lambda) - \frac{1}{2(f-q)}(\lambda - \sqrt{q}\sigma)^2} \right] + \right. \\ \left. + \int_{-\infty}^{+\infty} D\varphi \ln \left[\int_{-\infty}^{+\infty} d\mu(J) \exp\left[-\frac{1}{2}(2\hat{f} + \hat{q} + 2\beta\rho)J^2 + \sqrt{\hat{q}}J\varphi\right] \right] \right\} \quad (\text{C.44}) \end{aligned}$$

hierbij is de integratiemaat $d\mu(J)$ weer de oorspronkelijke maat die betrekking heeft op het minimalisatiedomein van de kostfunctie. Vermits we enkel geïnteresseerd zijn in de oplossingen van dit systeem in de hyperkubus zullen we in wat volgt ons daartoe beperken en dus nemen we $d\mu(J) = 1/2 \Theta(1 - J^2)$. De zadelpuntvergelijkingen van het precursor systeem in de hyperkubus volgen dan onmiddellijk. We zullen zoals voorheen afleiden naar $q, \hat{q}, (f - q)$ en in dit geval naar $(2\hat{f} + \hat{q} + 2\beta\rho)$. Dit gaat eenvoudig als we $f\hat{f} + \frac{1}{2}q\hat{q}$ schrijven als:

$$\begin{aligned} f\hat{f} + \frac{1}{2}q\hat{q} &= \frac{1}{2}(f - q)(2\hat{f} + \hat{q} + 2\beta\rho) + \\ &+ \frac{q}{2}(2\hat{f} + \hat{q} + 2\beta\rho) - (f - q)\beta\rho - \beta q\rho - \frac{\hat{q}}{2}(f - q) \quad (\text{C.45}) \end{aligned}$$

Voor het nemen van de limiet $\beta \rightarrow \infty$ voeren we zoals in sectie C.1 de volgende parameters in:

$$x = \beta(f - q) \quad (\text{C.46})$$

$$t = \frac{\sqrt{\hat{q}}}{2\hat{f} + \hat{q} + 2\beta\rho} \quad (\text{C.47})$$

$$y = (f - q)(2\hat{f} + \hat{q} + 2\beta\rho) \quad (\text{C.48})$$

C.5. De kostfunctie $E(J)$ met de perturbatieterm ρJ^2

Aldus vinden we de volgende zadelpuntvergelijkingen:

$$y - 2\rho x = -\alpha \int_{-\infty}^{+\infty} D\sigma \left[\frac{\partial \lambda_o}{\partial \sqrt{f}\sigma}(\sqrt{f}\sigma, x) - 1 \right], \quad (\text{C.49})$$

$$(ty)^2 = \alpha \int_{-\infty}^{+\infty} D\sigma \left[\lambda_o(\sqrt{f}\sigma, x) - \sqrt{f}\sigma \right]^2, \quad (\text{C.50})$$

$$y = \text{Erf} \left(\frac{1}{t\sqrt{2}} \right), \quad (\text{C.51})$$

$$f = 1 - y + yt^2 - \sqrt{\frac{2}{\pi}} t \exp\left[\frac{-1}{2t^2}\right] \quad (\text{C.52})$$

C. $V(\lambda)$ /MSB in de hyperkubus

Bijlage D

Over de convexiteit van kostfuncties

In deze appendix zal ik de convexiteit bewijzen van de gebruikte precursor kostfuncties op hun respectievelijk minimalisatiedomeinen. Dit impliceert dan de uniciteit van het minimum van de kostfunctie.

D.1 Definitie van convexiteit van een functie

Beschouw de functie $E(\mathbf{J})$ gedefinieerd op een domein $\mathcal{U}$ van een N -dimensionale ruimte. Beschouw twee vektoren $\mathbf{J}_1$ en $\mathbf{J}_2$ beiden uit het domein $\mathcal{U}$. De functie $E(\mathbf{J})$ heet dan convex als:

$$E(a\mathbf{J}_1 + (1-a)\mathbf{J}_2) \leq aE(\mathbf{J}_1) + (1-a)E(\mathbf{J}_2) \quad \forall a \in [0, 1] \quad (\text{D.1})$$

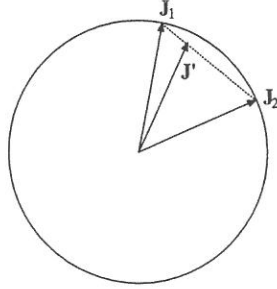
D.2 De kostfunctie bekomen door de potentiaal $V_{qo}(\lambda)$ op de hypersfeer

De potentiaal V_{qo} is gedefinieerd als:

$$V_{qo}(\lambda) \equiv \begin{cases} \frac{1}{\lambda - \kappa_b} & : \text{als } \lambda \geq \kappa_b \\ +\infty & : \text{als } \lambda < \kappa_b \end{cases} \quad (\text{D.2})$$

Bemerkt dat deze potentiaal voor $\lambda \geq \kappa_b$ voldoet aan de volgende twee ongelijkheden:

D. Over de convexiteit van kostfuncties



Figuur D.1: De vectoren $\mathbf{J}_1$, $\mathbf{J}_2$ en $\mathbf{J}'$ zoals beschreven in de tekst.

$$V_{qo}(a\lambda^1 + (1-a)\lambda^2) \leq aV_{qo}(\lambda^1) + (1-a)V_{qo}(\lambda^2) \quad \forall a \in [0, 1] \quad (\text{D.3})$$

en,

$$V_{qo}(\rho\lambda) \leq V_{qo}(\lambda) \quad \forall \rho \geq 1 \quad (\text{D.4})$$

Eigenschap (D.4) is triviaal en (D.3) kan m. b. v (D.2) herschreven worden tot:

$$\frac{(\lambda^1 - \kappa_b)(\lambda^2 - \kappa_b)}{(a\lambda^1 + (1-a)\lambda^2 - \kappa_b)(a(\lambda^2 - \lambda^1) + \lambda^1 - \kappa_b)} \leq 1 \quad (\text{D.5})$$

of nog:

$$\frac{1}{1 + \frac{a(\lambda^1 - \lambda^2) \cdot ((\lambda^1 - \kappa_b) - (\lambda^2 - \kappa_b))}{(\lambda^1 - \kappa_b) \cdot (\lambda^2 - \kappa_b)}} - \frac{a^2(\lambda^1 - \lambda^2)^2}{(\lambda^1 - \kappa_b) \cdot (\lambda^2 - \kappa_b)} \leq 1 \quad (\text{D.6})$$

Deze vergelijking is voldaan als:

$$a(\lambda^1 - \lambda^2) \cdot ((\lambda^1 - \kappa_b) - (\lambda^2 - \kappa_b)) \geq a^2(\lambda^1 - \lambda^2)^2, \quad (\text{D.7})$$

wat zowel voor $\lambda^1 > \lambda^2$ en $\lambda^2 > \lambda^1$ waar is.

Beschouw nu twee vectoren $\mathbf{J}_1$ en $\mathbf{J}_2$ met $\lambda_\mu^1 > \kappa_b$ en $\lambda_\mu^2 > \kappa_b$ en dit $\forall \mu (\mu : 1 \dots p)$. Neem de twee vectoren op de hypersfeer, dus $\mathbf{J}_1^2 = \mathbf{J}_2^2 = N$. Beschouw verder de vektor $\mathbf{J}' = a\mathbf{J}_1 + (1-a)\mathbf{J}_2$, waarbij $a \in [0, 1]$. De vektor $\mathbf{J}'$ ligt op de lijn die de twee vectoren $\mathbf{J}_1$ en $\mathbf{J}_2$ verbindt zoals geschetst in figuur (D.1).

Dan geldt er voor de kostfunctie $E(\mathbf{J}) = \sum_\mu V_{qo}(\lambda_\mu)$ met behulp van (D.3):

D.2. De kostfunctie bekomen door de potentiaal $V_{qo}(\lambda)$ op de hypersfeer

$$E(\mathbf{J}') \leq aE(\mathbf{J}_1) + (1 - a)E(\mathbf{J}_2). \quad (\text{D.8})$$

Dit bewijst de convexiteit in de hypersfeer. Renormaliseer de vektor $\mathbf{J}'$ zodat hij op de hypersfeer komt te liggen: $\mathbf{J} = \rho\mathbf{J}'$ met $\rho \geq 1$. Er volgt onmiddellijk met (D.4):

$$E(\mathbf{J}) \leq E(\mathbf{J}') \quad (\text{D.9})$$

zodat

$$E(\mathbf{J}) \leq aE(\mathbf{J}_1) + (1 - a)E(\mathbf{J}_2). \quad (\text{D.10})$$

Dit bewijst de convexiteit op de hypersfeer.

D.2.1 De kostfunctie bekomen door de potentiaal $V_{qo}(\lambda)$ in de hyperkubus

In dit geval moeten we de convexiteit bewijzen van de kostfunctie in de hyperkubus. Vermits de hyperkubus een convex domein is volstaat het aan te tonen dat $E(\mathbf{J})$ convex is. Dit kan eenvoudig bewezen worden als we aantonen dat de potentiaal $V_{qo}(\lambda)$ convex is. Dit is echter reeds aangetoond in vorige sectie vermits de potentiaal voldoet aan (D.3). Hiermee is de convexiteit in de hyperkubus bewezen.

De kostfunctie $E(\mathbf{J})$ met de perturbatieterm $\rho\mathbf{J}^2$

We zullen nu de convexiteit bewijzen in de hyperkubus als bij de kostfunctie de perturbatieterm $\rho\mathbf{J}^2$ toegevoegd is. Vermits de kostfunctie zonder de perturbatie convex is in de hyperkubus volstaat het te bewijzen dat:

$$\rho(a\mathbf{J}_1 + (1 - a)\mathbf{J}_2)^2 \leq a\rho\mathbf{J}_1^2 + (1 - a)\rho\mathbf{J}_2^2 \quad \forall a \in [0, 1], \quad (\text{D.11})$$

Uitwerken van bovenstaande uitdrukking levert:

$$-(1 - a)a\rho(\mathbf{J}_1^2 + \mathbf{J}_2^2) + 2(1 - a)\rho a\mathbf{J}_1 \cdot \mathbf{J}_2 \leq 0 \quad (\text{D.12})$$

Indien $\rho \geq 0$ is de eerste term steeds negatief daar $a \in [0, 1]$. De tweede term is dan altijd positief. Vermits er verder geldt dat

D. Over de convexiteit van kostfuncties

$$\mathbf{J}_1^2 + \mathbf{J}_2^2 \geq 2\mathbf{J}_1 \cdot \mathbf{J}_2, \quad (\text{D.13})$$

is de ongelijkheid geldig als $\rho \geq 0$.

Bijlage E

Formularium

E.1 Notaties

vektoren en matrices

Een N -dimensionale vektor wordt vet genoteerd, zoals: $\mathbf{S}$. De componenten van deze vektor worden aangeduid met sub-scripts $i, j, \dots$: S_i .

Het scalair product tussen twee vektoren $\mathbf{S}$ en $\mathbf{T}$ is gedefinieerd als:

$$\mathbf{S} \cdot \mathbf{T} \equiv \sum_j S_j T_j. \quad (\text{E.1})$$

Tevens wordt $\mathbf{S} \cdot \mathbf{S}$ ook wel genoteerd als $\mathbf{S}^2$.

Een matrix wordt genoteerd als $\{J_{ij}\}$. Het element op de i -de rij en j -de kolom is J_{ij} .

Gaussische integratiematen

De 1-dimensionale Gaussische integratiemaat:

$$Dz \equiv \frac{dz}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right]. \quad (\text{E.2})$$

Voor de 2-dimensionale Gaussische integratiemaat met correlatie γ :

$$D_\gamma(u, v) \equiv \frac{du dv}{2\pi\sqrt{1-\gamma^2}} \exp\left[-\frac{1}{2(1-\gamma^2)}(u^2 + v^2 - 2\gamma uv)\right]. \quad (\text{E.3})$$

E. Formularium

E.2 Enkele eenvoudige functies

De **sign**-functie is gegeven door:

$$\text{sign}(x) \equiv \begin{cases} -1 & : \text{ als } x < 0 \\ +1 & : \text{ als } x \geq 0. \end{cases} \quad (\text{E.4})$$

De **Θ** -functie is gedefinieerd als:

$$\Theta(x) \equiv \begin{cases} 0 & : \text{ als } x < 0 \\ +1 & : \text{ als } x \geq 0. \end{cases} \quad (\text{E.5})$$

De **Error**-functie $\text{Erf}(x)$ is gedefinieerd als:

$$\text{Erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x dt \exp[-t^2]. \quad (\text{E.6})$$

De **Gardner**-functie $H(x)$ is gedefinieerd als:

$$H(x) \equiv \int_x^{+\infty} Dz = \frac{1}{2} [1 - \text{Erf}(\frac{x}{\sqrt{2}})]. \quad (\text{E.7})$$

E.3 De Fouriertransformatie

De Fouriergetransformeerde $F(\omega)$ van een functie $f(x)$ is gedefinieerd als:

$$F(\omega) \equiv \int_{-\infty}^{+\infty} dx f(x) \exp[-i\omega x], \quad (\text{E.8})$$

met als inverse Fouriergetransformeerde:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} d\omega F(\omega) \exp[i\omega x]. \quad (\text{E.9})$$

E.4 Enkele identiteiten.

Gaussische integralen

$$\int_{-\infty}^{+\infty} dx \exp[-\frac{1}{2}ax^2 + bx] = \sqrt{\frac{2\pi}{a}} \exp[\frac{b^2}{2a}] \quad (a > 0). \quad (\text{E.10})$$

Afleidende naar b levert dan tevens ook:

E.4. Enkele identiteiten.

$$\int_{-\infty}^{+\infty} dx \exp\left[-\frac{1}{2}ax^2 + bx\right] = \sqrt{\frac{2\pi}{a}} \exp\left[\frac{b^2}{2a}\right] \left(\frac{b}{a}\right). \quad (\text{E.11})$$

Een speciaal geval van (E.10) is de Hubbard-Stratonovich formule:

$$\exp\left[\frac{p^2}{2}\right] = \int_{-\infty}^{+\infty} Dz \exp[pz]. \quad (\text{E.12})$$

met Dz zoals in (E.2). Verder is ook de fouriergetransformeerde van een gaussische functie vervat in (E.10):

$$\exp\left[\frac{-p^2}{2}\right] = \int_{-\infty}^{+\infty} Dz \exp[\pm i pz]. \quad (\text{E.13})$$

Nu volgen enkele Gaussische integralen met betrekking tot de integratiematen (E.2) en (E.3).

$$\int_{-\infty}^{+\infty} Dz = 1, \quad \int_{-\infty}^{+\infty} Dz z = 0, \quad \int_{-\infty}^{+\infty} Dz z^2 = 1, \quad (\text{E.14})$$

en verder

$$\int_{-\infty}^{+\infty} Dz z f(z) = \int_{-\infty}^{+\infty} Dz f'(z). \quad (\text{E.15})$$

Dit geldt alleen als $f'(z) \exp\left(\frac{-z^2}{2}\right) \rightarrow 0$ voor $z \rightarrow \pm\infty$ en kan eenvoudig getoond worden door partiële integratie.

Veralgemeningen van voorgaande integralen zijn:

$$\int_a^{+\infty} Dz z = \frac{1}{2\pi} \exp\left[-\frac{a^2}{2}\right], \quad (\text{E.16})$$

en

$$\int_a^{+\infty} Dz z^2 = \frac{1}{2\pi} \exp\left[-\frac{a^2}{2}\right] a + H(a), \quad (\text{E.17})$$

en

$$\int_a^{+\infty} Dz z f(z) = \int_a^{+\infty} Dz f'(z), \quad (\text{E.18})$$

waarbij de laatste uitdrukking weer eenvoudig kan getoond worden met partiële integratie en is geldig alleen als $f'(z) \exp\left(\frac{-z^2}{2}\right) \rightarrow 0$ voor $z \rightarrow +\infty$ en $z \rightarrow a$. Verder geldt er voor de twee-dimensionale integratiemaat:

E. Formularium

$$\int_{-\infty}^{+\infty} D_{\gamma}(u, v) f(u) = \int_{-\infty}^{+\infty} Du f(u), \quad (\text{E.19})$$

en

$$\int_{-\infty}^{+\infty} D_{\gamma}(u, v) u f(v) = \gamma \int_{-\infty}^{+\infty} Dv v f(v). \quad (\text{E.20})$$

De laatste twee eigenschappen zijn natuurlijk symmetrisch in u en v .

Representaties van Dirac-delta functies

Al de volgende uitdrukkingen zijn identisch met de Dirac-delta functie $\delta(x - x_o)$:

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} dy \exp[iy(x - x_o)], \quad (\text{E.21})$$

$$\lim_{\alpha \rightarrow +\infty} \sqrt{\frac{\alpha}{\pi}} \exp[-\alpha^2(x - x_o)^2], \quad (\text{E.22})$$

$$\lim_{\beta \rightarrow \infty} \frac{e^{-\beta(x-x_o)^2}}{\int_{-\infty}^{+\infty} dx e^{-\beta(x-x_o)^2}}. \quad (\text{E.23})$$

De laatste uitdrukking kan algemener geschreven worden als:

$$\lim_{\beta \rightarrow \infty} \frac{\Theta(a^2 - x^2) e^{-\beta(x-x_o)^2}}{\int_{-a}^{+a} dx e^{-\beta(x-x_o)^2}} = \begin{cases} \delta(x - x_o) & : \text{als } |x_o| \leq a \\ \delta(x - a) & : \text{als } x_o > a \\ \delta(x + a) & : \text{als } x_o < -a. \end{cases} \quad (\text{E.24})$$

Andere nuttige identiteiten

$$\begin{aligned} \frac{q}{2} y^2 + \frac{Q}{2} Y^2 + ryY &= \\ &= \frac{1}{4}(1 + \gamma)(\sqrt{q}y + \sqrt{Q}Y)^2 + \frac{1}{4}(1 - \gamma)(\sqrt{q}y - \sqrt{Q}Y)^2, \end{aligned} \quad (\text{E.25})$$

met $\gamma = \frac{r}{\sqrt{qQ}}$.

De Schwartz ongelijkheid:

E.4. Enkele identiteiten.

$$\left[\int dx f(x).g(x) \right]^2 \leq \left[\int dx f^2(x) \right] \cdot \left[\int dx g^2(x) \right], \quad (\text{E.26})$$

waarbij de gelijkheid geldt als $f(x) = Cg(x)$ met C een constante.

E. Formularium

Bibliografie

- [1] D. J. Amit, H. Gutfreund, and H. Sompolinsky. Spin-glass models of neural networks. *Phys. Rev. A*, 32(2):1007–1018, 1985.
- [2] J. K. Anlauf and M. Biehl. The AdaTron: an adaptive perceptron algorithm. *Europhys. Lett.*, 10:687–692, 1989.
- [3] G. J. Bex. *Tuning the Transfer Function: The Reversed Wedge and Beyond*. PhD thesis, LUC Diepenbeek/Belgium, 1996.
- [4] M. Bouten, J. Schietse, and C. Van den Broeck. Gradient descent learning in perceptrons: A review of its possibilities. *Phys. Rev. E*, 52:1958–1967, 1995.
- [5] B. Derrida, R. B. Griffiths, and A. Prügel-Bennett. Finite-size effects and bounds for perceptron models. *J. Phys. A:Math. Gen.*, 24:4907–4940, 1991.
- [6] S. Diederich and M. Opper. Learning of correlated patterns in spin-glass networks by local learning rules. *Phys. Rev. Lett.*, 58:949–952, 1987.
- [7] J. F. Fontanari and R. Meir. Evolving a learning algorithm for the binary perceptron. *Network*, 2:353–359, 1991.
- [8] B. M. Forrest. Content-addressability and learning in neural networks. *J. Phys. A:Math. Gen.*, 21:245–255, 1988.
- [9] E. Gardner. The space of interactions in neural network models. *J. Phys. A:Math. Gen.*, 21:257–270, 1988.
- [10] E. Gardner and B. Derrida. Optimal storage properties of neural networks. *J. Phys. A:Math. Gen.*, 21:271–284, 1988.

BIBLIOGRAPHIE

- [11] M. Griniasty and H. Gutfreund. Learning and retrieval in attractor neural networks above saturation. *J. Phys. A:Math. Gen.*, 24:715–734, 1991.
- [12] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA*, 79:2554–2558, 1982.
- [13] H. Horner. Dynamics of learning for the binary perceptron problem. *Z. Phys. B*, 86:291, 1992.
- [14] J. Iwanski, J. Schietse, and M. Bouten. Replica symetrie breaking in a diluted network with binary couplings. *Phys. Rev. E*, 52:888–894, 1995.
- [15] I. Kanter and H. Sompolinsky. Associative recall of memory without errors. *Phys. Rev. A*, 35:380–392, 1987.
- [16] T. B. Kepler and L. F. Abbott. Domains of attraction in neural networks. *J. Phys. France*, 49:1657–1662, 1988.
- [17] O. Kinouchi and N. Caticha. Learning algorithm that gives the bayes generalization limit for perceptrons. *Phys. Rev. E*, 54:R54–R57, 1996.
- [18] H. M. Köhler. Adaptive genetic algorithm for the binary perceptron problem. *J. Phys. A:Math. Gen.*, 23:L1265–L1271, 1990.
- [19] W. Krauth and M. Mézard. Learning algorithms with optimal stability in neural networks. *J. Phys. A:Math. Gen.*, 20:L745–L752, 1987.
- [20] W. Krauth and M. Mézard. Storage capacity of memory networks with binary couplings. *J. Phys. France*, 50:3057–3066, 1989.
- [21] W. Krauth, J.-P. Nadal, and M. Mézard. The roles of stability and symmetry in the dynamics of neural networks. *J. Phys. A:Math. Gen.*, 21:2995–3011, 1988.
- [22] W. Krauth and M. Oppen. Critical storage capacity of the $j=1$ neural network. *J. Phys. A:Math. Gen.*, 22:L519–L523, 1989.
- [23] G. Milde and S. Kobe. An exact learning algorithm for autoassociative neural networks with binary couplings. *J. Phys. A:Math. Gen.*, 30:2349–2352, 1997.

BIBLIOGRAFIE

- [24] W. Nadler and W. Fink. Finite size scaling in neural networks. *Phys. Rev. Lett.*, 78:555–558, 1997.
- [25] M. Opper, J. Kleinz, H. Köhler, and W. Kinzel. Basins of attraction near the critical storage capacity for neural networks with constant stabilities. *J. Phys. A:Math. Gen.*, 22:L407–L411, 1989.
- [26] R. W. Penney and D. Sherrington. The weight-space of the binary perceptron. *J. Phys. A:Math. Gen.*, 26:6173–6185, 1993.
- [27] C. J. Pérez Vicente, J. Carrabina, and E. Valderrama. Study of a learning algorithm for neural networks with discrete synaptic couplings. *Network*, 3:165–176, 1992.
- [28] L. Pitt and L. G. Valiant. Computational limitations on learning from examples. *Journal of ACM*, 35:965–984, 1988.
- [29] G. Pöppel and U. Krey. Dynamical learning process for recognition of correlated patterns in symmetric spin glass models. *Europhys. Lett.*, 4:979, 1987.
- [30] E. M. Reingold, J. Nievergelt, and N. Deo. *Combinatorial Algorithms: Theory and Practice*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1977.
- [31] J. Schietse, M. Bouten, and C. Van den Broeck. Training binary perceptrons by clipping. *Europhys. Lett.*, 32:279–284, 1995.
- [32] T.L.H. Watkin. Optimal learning with a neural network. *Europhys. Lett.*, 21:871–876, 1993.
- [33] A. C. Wilson. The molecular basis of evolution. *Sc. Am.*, 253:148, 1985.
- [34] K. Y. M. Wong, A. Rau, and D. Sherrington. Weight space organization of optimized neural networks. *Europhys. Lett.*, 19:559–564, 1992.

BIBLIOGRAFIE

Publikatielijst

- L. Reimers, M. Bouten and B. Van Rompaey, "Learning Strategy for the Binary Perceptron", J. Phys. A.: Math. Gen. **29** (1996) 6247–6252.
- B. Van Rompaey, "Precursor Networks for Training the Binary Perceptron", Proceedings of ESANN'97 — Brugge, 16, 17–18 april, Ed. M. Verleysen, D. Facto Publications (p175–180).
- M. Bouten, L. Reimers and B. Van Rompaey, "Learning in the Hypercube: A Stepping Stone to the Binary Perceptron", Phys. Rev. E **58** (1998) 2378–2385.
- K. Meykens, B. Van Rompaey and H. Janssen, "Dispersion in Acoustic Waveguides: A Teaching Laboratory Experiment", Am. J. Phys. **67**(5) (1999) 400–406
- M. Copelli, M. Bouten, C. Van den Broeck and B. Van Rompaey, "On the Center of Mass of Ising Vectors", submitted to Europhys. Lett.
- B. Van Rompaey, "Precursors and Binary Trees for Learning the Binary Perceptron", submitted to Phys. Rev. E.

