



室蘭工業大学

学術資源アーカイブ

Muroran Institute of Technology Academic Resources Archive



初期視覚野における非線型ニューロン応答特性の分析手法及び学習モデルに関する研究

メタデータ	言語: jpn 出版者: 公開日: 2013-11-15 キーワード (Ja): キーワード (En): 作成者: 横山, 裕樹 メールアドレス: 所属:
URL	https://doi.org/10.15118/00005097

初期視覚野における非線型ニューロン応答特性
の分析手法及び学習モデルに関する研究

博士論文

横山 裕樹

室蘭工業大学大学院工学研究科生産情報システム工学専攻

平成 24 年度

目次

第1章	序論	7
1.1	脳機能の生理学的知見	7
1.1.1	ニューロンの応答特性	7
1.1.2	シナプスの可塑性	9
1.2	視覚系	11
1.2.1	網膜と視床	11
1.2.2	視覚皮質	12
1.3	視覚皮質の計算理論	14
1.3.1	計算論的神経科学	14
1.3.2	ニューロン応答のモデル	16
1.3.3	大脳皮質の教師なし学習	17
1.3.4	生成モデル	22
1.3.5	階層モデル	23
1.4	初期視覚ニューロン応答の測定手法	24
1.4.1	逆相関法	25
1.4.2	逆相関法の理論	26
1.5	カーネル法	27
1.5.1	カーネル関数の条件	29
1.5.2	カーネル関数の構成	30
1.5.3	カーネル関数の例	30
1.6	本研究の目的	31
第2章	初期視覚野の非線型ニューロン応答の分析手法	33
2.1	目的	33
2.2	ニューロン応答の区分線形近似	34
2.3	逆相関法による局所線形カーネルの推定	36
2.4	数値実験	37
2.4.1	複雑型細胞モデル	38
2.4.2	方位交差抑制モデル	41

2.5	まとめ	44
第3章	階層ベイズモデルによる組み合わせ表現の獲得	47
3.1	目的	47
3.2	モデル	48
3.3	勾配法による MAP 推定	51
3.4	改良モデル	55
3.5	変分ベイズの適用	56
3.6	まとめ	59
第4章	生成モデルへのカーネル法の導入	61
4.1	目的	61
4.2	モデル	62
4.2.1	非負性	62
4.2.2	構造	63
4.3	カーネル関数	64
4.3.1	カーネル関数の条件	64
4.3.2	カーネル関数と特徴空間の関係	65
4.3.3	不定値カーネル	66
4.3.4	カーネル関数の微分	68
4.4	最適化問題の収束性	69
4.5	数値実験	71
4.5.1	方法	71
4.5.2	基底	72
4.5.3	スパースネス	72
4.5.4	方位選択性	72
4.6	考察	75
4.6.1	ニューロン間の相互作用	76
4.6.2	基底ベクトルの正規化	79
4.6.3	多層化	79
4.7	幾何学的考察	80
4.7.1	テンソルの表記法	81
4.7.2	入力画像と基底画像の空間	81
4.7.3	予測画像の空間	84
4.7.4	応答特性	85
4.7.5	具体例	88
4.8	まとめ	90

	5
第 5 章 結論	93
参考文献	97
研究業績	109

第1章 序論

近年の計測技術の進歩により，脳の視覚機能について様々な知見が得られている．しかしながら，非線型で確率的に応答するニューロンの特性を適切に記述し，実験データからそのパラメータを推定することは未だ困難な課題である．また，その特性が脳内の神経回路において計算論的にどのような役割を持つかについても，様々なモデルが提案されているが，未だ未解決な問題も存在する．以上のような背景の下で，本研究は大きく分けて二つの課題に取り組んでいる．

一つは初期視覚野の応答特性を記述するモデルと，それに基づく電気生理的な実験手法の提案である．ニューロン応答は多入力非線型関数と，ポアソンのスパイク生成過程でモデル化することができる [1]．強い非線型性を持つ多入力関数を表現し，分析することは困難である．本研究では，非線型関数を区分関数として表現し，各領域について基底関数の線型和として推定する手法を提案する．

もう一つは初期視覚野の計算モデルに関する研究である．初期視覚領野における受容野の自己組織的獲得や，領野間の双方向性結合の役割は，生成モデルの枠組みで説明されることが多い．特に第一次視覚野が表現している線型基底の自己組織的獲得については，生成モデルでよく再現することができる．一方で，視覚野のニューロンは expansive な非線型性を持つことが知られている [2]．従ってこの非線型性の役割についても理解する必要がある．本研究では非負制約のある生成モデルにカーネル法を導入し，ニューロン応答の非線型性が応答特性や受容野獲得に及ぼす影響について検証する．

本章では，まず本研究の背景となる脳機能に関する知見と数理モデルについて概説する．次にこの背景を踏まえた本研究の位置づけと，各章における具体的なアプローチについて述べる．

1.1 脳機能の生理学的知見

1.1.1 ニューロンの応答特性

脳を含む神経系は，多数の神経細胞からなる複雑なネットワークを形成することで，その高度な機能を実現している．この神経細胞は，神経系の機能単位とい

う意味でニューロン (neuron) と呼ばれている。ニューロンは細胞体の他に樹状突起と軸索という特異な構造を持ち、他のニューロンとシナプス結合を作ることによって情報の伝達を行っている。機能的には、樹状突起が入力を受け、軸索が出力を担っている。多くの場合、細胞体から伸びる数本の樹状突起が枝分かれし、数百から数千のシナプスを形成しており、軸索は一本である。このことから、ニューロンは機能的に多入力一出力の系であるといえる。

ニューロンの細胞膜上には、イオンチャネルと呼ばれる特定のイオンを透過させるタンパク質が分布しており、これらが条件に応じて開閉することで細胞内外の電位差 (膜電位) を変化させている。入力が無い状況では、カリウム漏洩チャネルによるカリウムイオンの流出が支配的であり、これによって膜電位は、細胞外を基準とすると約 -70mV に保たれている。これは静止電位と呼ばれる。シナプスからの興奮性入力によって脱分極が起こると、電位依存性のナトリウムチャネルが開き、ナトリウムイオンが流入してさらに膜電位が上昇する。この正帰還による脱分極がカリウム漏洩による電流を上回ると、活動電位と呼ばれるスパイク状の電位が発生する。電位がピークの $+30\text{mV}$ 程度に達すると、ナトリウムチャネルが閉じるとともに電位依存性のカリウムチャネルが開き、カリウムイオンの流出によって再分極が起こる。その後ナトリウムチャネルが不活性状態となるため、活動電位を発生させることができない不応期と呼ばれる数 ms の期間を経て通常の状態に戻る。一方で、シナプスからの興奮性入力による脱分極がカリウム漏洩による電流を上回るほどではない場合、活動電位は発生せず、入力が弱まれば静止電位に戻る。活動電位が軸索に伝わると、上記の反応が連鎖的に起こり、電気信号は軸索の先端まで減衰することなく伝わる。また、不応期があるため活動電位の逆行は起こらない。

シナプスには化学シナプスと電気シナプスがある。化学シナプスは軸索終末などのシナプス前要素、樹状突起などのシナプス後要素と、それらの間のシナプス間隙からなる構造である。前要素に到達した電気信号によってカルシウムチャネルからカルシウムイオンが流入すると、前要素に配置されたシナプス小胞から神経伝達物質がシナプス間隙に分泌される。後要素には受容体が存在し、これに神経伝達物質が結合することで、情報は再び電気信号に変換される。化学シナプスはさらに、後要素を脱分極させる興奮性シナプスと、過分極させる抑制性シナプスに分けられ、これは軸索終末が属するニューロンの種類に依存する。一方、電気シナプスは、膜電位を直接次のニューロンへ伝える構造で、系統学的には化学シナプスより古いものであるといわれている。

1.1.2 シナプスの可塑性

脳の非常に重要な機能の一つに学習がある。学習とは、ある知覚または行動などを一度または複数回経験することによって、知覚の感度が向上したり、より効率的な運動を行えるようになることを指す。これを実現するためには、脳の構造に何らかの可塑性が必要である。

視覚野における可塑性に関する知見は、HubelとWieselの仔ネコを用いた実験に始まる [3]。彼らは、仔ネコの片目を一定期間閉じると、皮質ニューロンがその遮蔽した目からの刺激に反応しなくなることを見いだした。また、この可塑性は臨界期と呼ばれる生後発達の一時期によくみられ、視覚野の眼優位コラムと呼ばれる回路構造の変化がみられた [4]。

脳に可塑性を与えるメカニズムの最も大きな要素は、シナプス可塑性にあると考えられている。これは、ニューロン間の結合強度が変化することで神経回路の構造が変化し、それによって脳で行われる情報処理も変化すると考えられるためである。Hebb[5]は、シナプス可塑性と脳の学習とを結びつける、次の重要な仮説を提唱している：

細胞 A の軸索が細胞 B を発火させるのに十分近くにあり、繰り返しあるいは絶え間なくその発火に参加するとき、いくつかの成長過程あるいは代謝変化が一方あるいは両方の細胞に起こり、細胞 B を発火させる細胞の 1 つとして細胞 A の効率が增加する。

この仮説は Hebb 則と呼ばれている。

シナプスの伝達効率は、その前後のニューロンの活動強度に応じて時々刻々と変化するが、特に長期増強 (long-term potentiation; LTP) と長期抑圧 (long-term depression; LTD) が、学習という側面において重要であると考えられる。LTP は、Bliss と Lomo[6] により海馬で発見された。彼らはウサギの海馬歯状回のニューロンに 100Hz の高頻度刺激を短時間与え、その後シナプスの伝達効率が持続的に増強することを発見した。LTP はその後、大脳皮質を含む様々な領域においても報告されている [7, 8, 9]。一方 Dudek と Bear[10] は海馬において、2Hz の低頻度刺激を長時間与えることで、対象のシナプスに LTD が生じることを発見した。

LTP と LTD は、スライス標本を用いた *in vitro* の実験では容易に誘発されることから、様々な研究が行われた。Bienenstock, Munro と Cooper は、刺激の強度に応じて生じる増強と抑圧の二つの現象を統一的に説明するモデルを提唱した [11]。このモデルは提唱者の頭文字をとって BMC モデルと呼ばれている。彼らは時刻 t におけるシナプスの結合加重 $m(t)$ が一時遅れ

$$\dot{m}(t) = \phi(c(t))d(t) - \epsilon m(t) \quad (1.1)$$

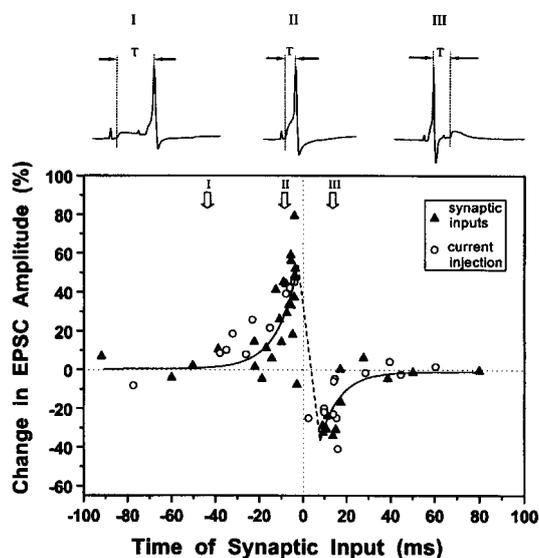


図 1.1: スパイクタイミング依存性シナプス可塑性 (STDP). [12] より転載.

に従うと仮定した. ここで, $d(t), c(t)$ はそれぞれシナプス前細胞とシナプス後細胞の応答を表す. $\phi(c)$ は c によって増減する m の増加率を示している. $\phi(c)$ は c がある閾値 θ を超えれば m が増加し, 下回れば m が減少するような曲線を描く. これはあるニューロンの活動が複数のニューロンを活動させたとき, 一部の結合が増強される一方で, 別の一部の結合は抑圧され得ることを意味している. また彼らは, θ 自体がより長期的な活動の履歴に応じて変化するとしている.

シナプス可塑性は刺激の強度だけでなく, タイミングにも依存することが報告されている [13]. この現象はニューロンの種類によって異なるが, 興奮性ニューロン間の結合では図 1.1 のような結果が得られている [14, 15, 12]. 横軸は長期刺激中にシナプス前細胞によって生じた EPSP とシナプス後細胞の活動電位の時間差で, 左は前者が先行した場合, 右は後者が先行した場合に相当する. 縦軸は長期刺激後にシナプスが増強または抑圧された割合を表している. このグラフより興奮性-興奮性の場合, 前細胞からの入力の後細胞の活動電位にわずかに先行した場合にシナプスが増強され, 逆の場合には抑圧されることがわかる.

Hebb の仮説は, 細胞 A と細胞 B の活動の因果関係について言及している. すなわち, 両者がただ同時に発火すればシナプスが増強されるのではなく, 細胞 A の活動が細胞 B の活動の要因の一つでなければならないことを明確に述べている. STDP という現象は, シナプス前細胞からの入力後細胞の発火に先行しているときのみ結合加重が増強されるという点において, この因果関係を支持

しているといえる [16]. このことから、後細胞の発火にそれほど寄与しなかったシナプスが減衰するという特性を除けば、Hebb の仮説は正しいといえる. 一方で、Hebb 則に従わないシナプス可塑性の法則も知られており、特に反 Hebb 則は重要な研究対象となっている. Hebb 則は統計学習と深い関わりがあり、後述するように様々な学習モデルの基礎となっている.

1.2 視覚系

1.2.1 網膜と視床

視覚情報の媒体は光であるが、神経情報のそれは前述の通り、基本的に電気信号である. このため視覚系が最初に行うのは、光から電気信号への変換である. 光源から放射された光は、直接または物体で反射することによって眼球へ入射し、水晶体で屈折した後、網膜へ達する. 網膜には錐体 (cone)、杆体 (rod) と呼ばれる視細胞が多数存在する. これらは視物質と呼ばれるタンパク質を持っており、これに光が当たることで緩やかなイオン電流を発生する. 錐体は視物質の分光特性の違いにより、ヒトの場合さらに三種類 (L, M, S) に分けられる. これら三種の錐体の応答を組み合わせることで、色が三次元的に表現される. 一方、杆体の視物質は一種類であるため色を区別できないが、感度が高く暗所で機能する. これら視細胞が網膜上の随所に配置され、光刺激に応じて活動することで、外界の三次元世界の投影像を表現する. 網膜の各所で電気信号に変換された視覚情報は、網膜の局所回路を経て神経節細胞でまとめられ、視床の外側膝状体 (lateral geniculate nucleus; LGN) へ送られる. 網膜の多くの細胞は光刺激に対して緩やかな電位変化を呈するが、神経節細胞の段階で、活動電位が視覚情報の媒体となる.

あるニューロンの入力となり得る視野上の範囲を、そのニューロンの受容野 (receptive field; RF) という. 受容野は円形とは限らず、特殊な形状を持つ場合がある. また、光刺激がそのニューロンを興奮させる点だけでなく、抑制する点が存在することもある. このため受容野という用語は、単に範囲だけでなく、ニューロンの機能的な、特に線型フィルタとしての側面を指して用いられる.

網膜神経節と LGN では、図 1.2(a) のような中心周辺拮抗型の受容野がみられる. 同心円で区切られた領域の + と - は、光刺激が興奮的に働くオン領域と抑制的に働くオフ領域を示している. このような受容野は、オン領域のみに光刺激を呈示したときに最も強く応答し、受容野全体に一樣な光刺激を呈示すると、スパイクの頻度が低下する. このことから、網膜神経節と LGN では主に、コントラストやエッジの検出を行っていると考えられる. 中心周辺拮抗型の受容野を線型フィル

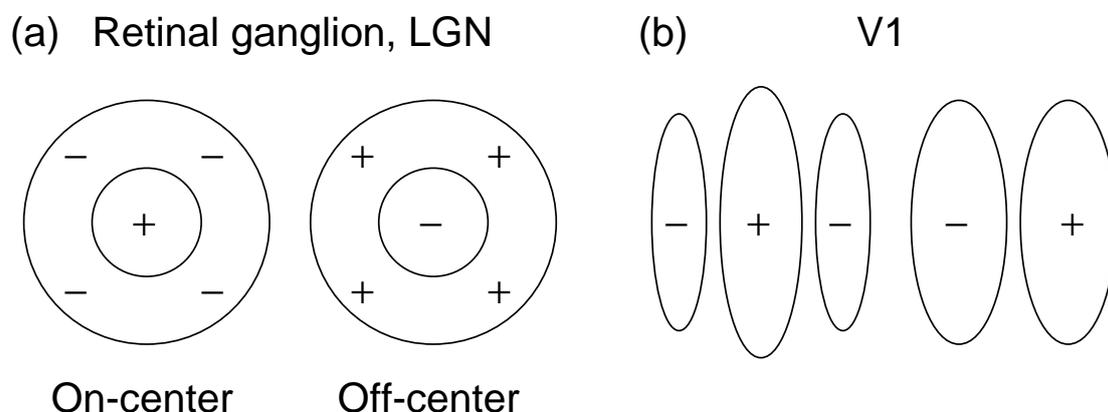


図 1.2: (a) 網膜神経節や外側膝状体で見られる on-center 型受容野と off-center 型受容野. (b) 第一次視覚野で見られる Gabor 型受容野. 正弦波の方位や周波数, 位相によって様々な形状を持つ.

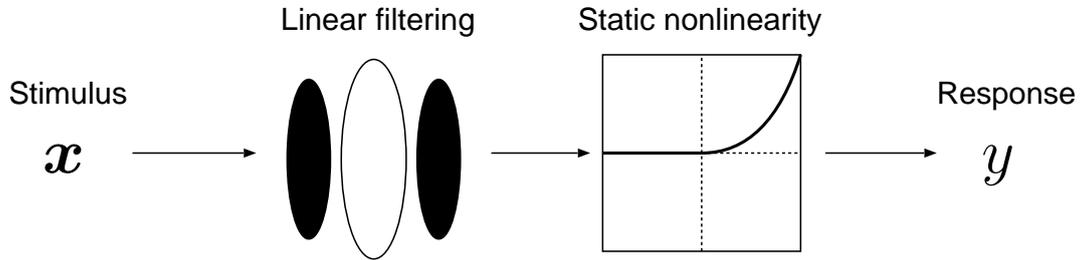
タとしてみるならば, これはラプラス作用素とローパスフィルタを組み合わせたバンドパスフィルタに相当する.

1.2.2 視覚皮質

LGN から大脳皮質への遠心性経路は, まず後頭葉の第一次視覚野 (V1) に到達する. その後の視覚経路は, V2, V4 を経由して下側頭葉 (IT) 皮質へ向かう腹側経路と, V2, V3, MT (middle temporal) 野, MST (medial superior temporal) 野を通過して頭頂葉へ至る背側経路の二つに大きく分かれる [17]. 前者は網膜像から物体の形状や色を抽出し, いま何を見ているのかを認識する機能に関与していると考えられている. 一方, 後者は物体の位置や動きの把握に関係していると言われている. しかし, それらが具体的にどのように実現されているのか, 二つの経路の計算結果はどのように統合されるのか, また計算過程でこれらが分離される意義についてはよくわかっていない.

以上のように, 大脳皮質は LGN 以前の部位よりも高度な視覚情報処理を行っている. その実現のため, 皮質ニューロンの受容野はより大きく複雑な形状を持っている. これを最初に明らかにしたのは Hubel と Wiesel [18, 19] である. 彼らはネコやサルに棒状の光刺激を呈示し, 微小電極により V1 ニューロンの電氣的活動を調べることにより, これらが特定の線分方位に選択的に応答することを発見した. この方位選択性を持つ V1 ニューロンの受容野は図 1.2(b), 正弦波を Gauss 窓で局

(a) Simple cell (linear-nonlinear model)



(b) Complex cell (energy model)

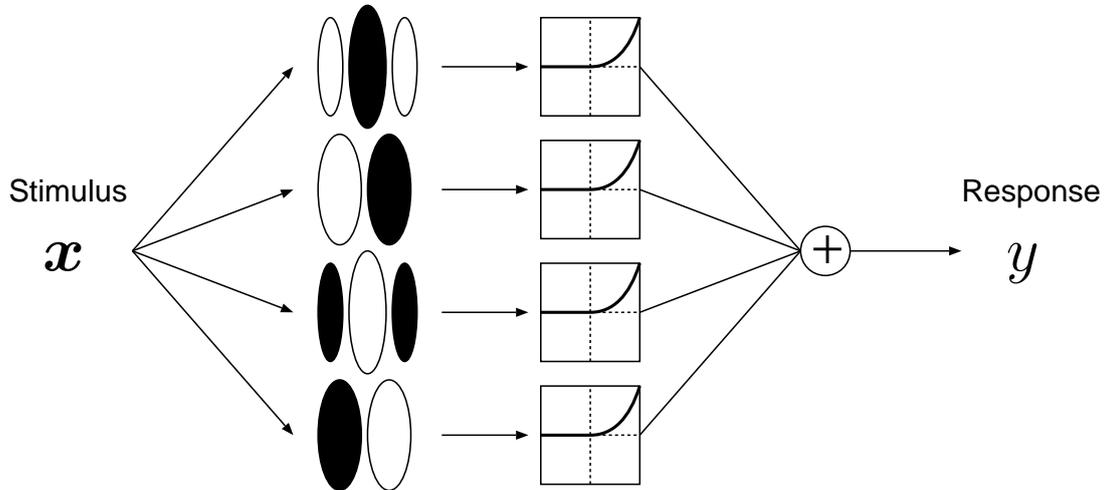


図 1.3: 単純型細胞の LN モデル (a) と複雑型細胞のエネルギーモデル (b).

在化させた Gabor 関数

$$G(\mathbf{p}) = \exp\left(-\frac{1}{2}(\mathbf{p} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{p} - \boldsymbol{\mu})\right) \cos(2\pi \mathbf{f}^T \mathbf{p} + \phi) \quad (1.2)$$

で近似することができる. $\boldsymbol{\mu} \in \mathbb{R}^2$ は受容野の中心を表す. 受容野の中心を原点とすれば $\boldsymbol{\mu} = \mathbf{0}$ である. $\boldsymbol{\Sigma} \in \mathbb{R}^{2 \times 2}$ は正定値対称行列で, ガウス窓の幅を表す. $\boldsymbol{\Sigma}$ の固有値は, 対応する固有ベクトル方向の幅の二乗となっている. $\mathbf{f} = (f_x, f_y)^T$ は正弦波の x, y 方向の空間周波数を表し, ϕ は空間位相を表す. V1 ニューロンは特定方向への運動に選択性を持つものも発見されており (例えば [20]), その受容野は上式に時間軸を追加した三次元の Gabor 関数でモデル化される.

V1 ニューロンは単純型細胞と複雑型細胞の二種類に大別される. 単純型細胞は, 大まかには図 1.3(a) のように, 上記の Gabor 型受容野による線型フィルタリング

と時空間軸に対して static な非線型性からなる linear-nonlinear (LN) モデルで説明することができ、定性的には線分の位置に依存する線型フィルタとして振る舞う。一方、複雑型細胞は方位選択性を持ちながら空間位相には依存しない、非線形な特性を持つ。このため受容野の各点にオン、オフを対応付けることはできない。この非線型性を再現するモデルは図 1.3(b) のエネルギーモデル [21] が有名である。このモデルでは、複雑型細胞は四つの機能的サブユニットで構成される。各サブユニットは位相のずれた単純型細胞の LN モデルと等価である。入力線分の方位が最適方位に近ければ、様々な線分位置に対して常に一つ以上のサブユニットが応答し、その和を取ることで全体として一定の強度で応答することができる。

このような特性から、単純型細胞の集団による網膜像の表現はフーリエ変換（またはウェーブレット変換）に相当するといえる。また、この対応に従うならば、複雑型細胞による表現はパワースペクトルに相当する。V1 には色や両眼視差に選択性を持つニューロンも存在するが、基本的には網膜像の時空間的な周波数分析を行っているといえる。

V2 では二つの線分の組み合わせに選択性を持つニューロンが発見されている [22]。これらは V1 ニューロンの組み合わせで再現できることが示唆されるが [23]、より複雑なパターンに選択性を持つニューロンも発見されており [24]、統一的な説明は得られていない。さらに V4 では閉曲線の曲率 [25, 26] に選択的なニューロンが報告されており、腹側経路の終着となる IT では、図形の特徴的な組み合わせ [27, 28] や顔 [29, 30] など、物体の記述に直接関わる重要な情報を符号化している可能性が示唆される。

1.3 視覚皮質の計算理論

1.3.1 計算論的神経科学

これまでは主にニューロンレベルの生理学的知見について述べてきたが、ヒトの脳は百億個のニューロンとその千から一万倍のシナプス結合で構成される複雑な系であり、ニューロンレベルの知見をいくら集めても脳を完全に理解したことはならない。そのためより細かい分子生物学的なよりミクロなメカニズムや、逆にモジュール単位や個体レベルの行動、さらには社会的行動といったマクロな知見も必要であり、実際様々なレベルで脳機能の研究が行われている。

脳科学の階層として考えられるのは、このようなシステムの構成要素としての階層だけではない。Marr は、脳を完全に理解したと言うためには、以下の三つの水準での理解が必要であると主張した [31]。

表 1.1: 脳の理解に必要な三つの水準 [31]. 本文では上から順に第一, 第二, 第三の水準と呼んでいる.

計算理論 (computational theory)	計算の目的は何か, なぜそれが適切なのか, そしてその実行可能な方略の理論は何か.
表現とアルゴリズム (representation and algorithm)	この計算理論はどのようにして実現することができるか. 特に入力と出力の表現は何か, そしてその変換のためのアルゴリズムは何か.
ハードウェアによる実現 (hardware implementation)	表現とアルゴリズムがどのようにして物理的に実現されるか.

第一の水準は最も抽象的な計算理論である. この水準では, 情報処理系が何を計算するのかを示すために, その動作をある情報から別の情報への写像として特徴付ける. そして, なぜその計算目標が課題の遂行に対して適切なのかを説明する. 例えば商店の勘定台にあるキャッシュレジスタは, 実質的には商品の価格の加算を実行しているのであり, これは購入された個々の商品の価格から最終的な勘定を得るために必要であると思われる手続きの制約条件が, 整数における加算の代数的規則と一致するからである.

第二の水準では, この計算理論をどのように実現するのかを示すために, 入力と出力の表現と, 入力を出力に変換するために実行されるアルゴリズムを記述する. キャッシュレジスタの例では, 入出力の表現は自然数であり, 桁の繰り上がりを含む加法の具体的手順がアルゴリズムに相当する.

第三の水準は表現とアルゴリズムが物理的にどうやって実現されるかを記述する. 機械式のキャッシュレジスタは, 歯車の角度などで数値を表現し, これを回転させることによって演算を行う. また電気式の場合は, コンデンサや半導体内部のキャリアの状態によって数値を表現し, 各素子の電圧の上げ下げによって演算を行う.

これら三つの水準は, 互いに依存し合いながらも, ある程度の独立性を保って存在する概念である. 近年では, コンピュータのプロセッサをトランジスタの密結合によって高速化することは, 量子論的な制約により限界に達してきている. このため, プロセッサの進歩はクロック数の増加からコア数の増加という形に変わってきており, 従来の逐次的なアルゴリズムではプロセッサの性能を十分に生かすことができなくなってきた. これは, ハードウェアの構造の変化がアルゴリズムの善し悪しに影響を与える例である. しかしこの変化によって, あるソフトウェアが何を何のために計算するのかという, 第一水準の問いに対する答えが変わることはない.

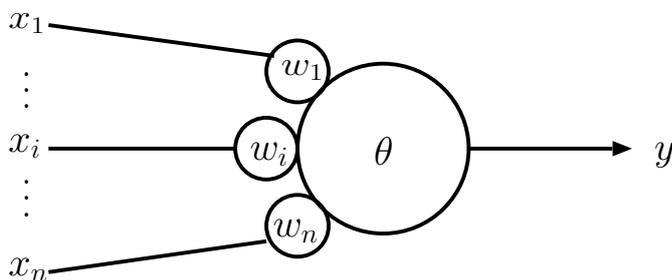


図 1.4: McCulloch-Pitts の閾値モデル.

川人は計算論的神経科学を以下のように定義している [32] :

脳機能を、その機能を脳と同じ方法で実現できる計算機のプログラムあるいは人工的な機械を作れる程度に、深く本質的に理解することを目指すアプローチを計算論的神経科学と呼ぶ。

ただし彼は、これはあくまで、「脳の理解」のレベルを客観的に定義するための制限であって、計算論的神経科学の真の目的は人工物をつくる工学ではなく脳を理解する理学であると断っている。脳の計算をアルゴリズムのレベルではわかっているとしても、それが何を計算しているのかを理解しなければ、その計算を全く異なるハードウェア基盤を持つコンピュータで実現することはできない。こういった意味で上の定義は、Marr の第一水準の理解が計算論的神経科学において非常に重要であることをよく表している。

1.3.2 ニューロン応答のモデル

神経回路の機能を計算論的に理解するためには、まずニューロンの電気的な活動を何らかの方法で記述する必要がある。しかし以下に示すように、ニューロンは電气的にみれば多数の電流源からなる複雑な系である。ニューロン自体の活動のメカニズムを探るためには、その構成要素であるイオンチャネルや受容体の特性を忠実に再現するモデル（例えば Hodgkin-Huxley モデル [33]、コンパートメントモデル [34] など）を構築して、それらの相互作用について分析する必要がある。一方、多数のニューロンからなる神経回路の機能を探るためには、個々のニューロン特性については本質を失わない程度に簡略化し、それらの相互作用に焦点を置かなければ本質に迫る結果は得られない。

McCulloch と Pitts[35] は，上で述べたイオンチャネルの競合を，シナプス入力
の合計に対する閾値処理とみて，以下の閾値モデルを提案した（図 1.4）．

$$y(t) = s \left(\sum_i w_i x_i(t) - \theta \right), s(u) = \begin{cases} 0 & (u < 0) \\ 1 & (\text{otherwise}) \end{cases} \quad (1.3)$$

ここで x_i と w_i はそれぞれシナプス i に対するシナプス前細胞の応答と結合加重， θ
は閾値を表す．彼らは， $\mathbf{w} = (w_1, \dots, w_m)^T$ （右上付きの T は行列の転置を表す）
と θ を適切に設定することによって AND や OR などの論理演算が実現できること
を示し，その後の計算機科学，人工知能などの研究に大きな影響を与えた．しかし閾値モデルは微分ができないという問題がある．ニューロンモデルの微分可能性は，勾配に基づいた学習に必要となるため重要である．そこで，ニューロンの発火頻度を出力と考える，以下の発火頻度モデルが提案されている．

$$y(t) = f \left(\sum_i w_i x_i(t) \right), f(u) = \frac{1}{1 + e^{-u+\theta}} \quad (1.4)$$

ここで θ は，発火頻度が最小値と最大値の間になる u の値を定める定数である．
この発火頻度モデルは，以下で参照する学習モデルを含む，多くのモデルの基礎
となっている．

1.3.3 大脳皮質の教師なし学習

これまで，LTP，LTD などのシナプス可塑性について，主に生理学的知見から
述べてきた．最近ではシナプス可塑性の分子生物学的機構については詳しく研究
されているが，その計算論的役割についても理解する必要がある．シナプス可塑
性によって実現されていると考えられている学習について計算論的説明を与える
枠組みは，大きく分けて三つある [36]．一つは，各入力信号に対して出力すべき正
しい信号（教師信号）が与えられ，実際に出力した信号と教師信号との誤差が小
さくなるように学習する，教師あり学習である．これは例えば，楽器の演奏の仕
方を教わってできるようになることや，手本に従って漢字の書き順を覚えること
に相当する．次に，教師信号は与えられないが，出力に対する評価（報酬）が与え
られ，これを最大化するように出力を選択して行く強化学習がある．例えば，将
棋や囲碁などのゲームにおいては，繰り返し対局を行って勝ち負けを経験するこ
とで，特に教えられなくても良い手を打つことができるようになる．これら二つ
の学習はそれぞれ主に小脳，大脳基底核で行われていると考えられている [37, 38]．

これに対して，視覚野を含む感覚系の皮質領野で行われていると考えられてい
る学習の枠組みは，教師なし学習と呼ばれる．教師なし学習では，望ましい出力

も、出力の善し悪しも与えられないため、なんらかの基準に従って入力を入力へ変換する処理過程を学習しなければならない。その基準として考えられる原理に情報量最大化がある [39, 40, 41]。脳には多数のニューロンが存在するとはいえ、その数は有限であり、またニューロンの電気的活動にはエネルギーを要する。たとえ脳が食餌の確保や危険の回避のために優れた情報処理を行っていたとしても、それが膨大なエネルギーを浪費してしまつては本末転倒である。従つて、脳は限られた資源を利用して効率的に情報処理を行う必要がある。そのためには、出力の冗長性を少なくすることが重要である [42, 43]。

いま多入力多出力の情報処理系を考え、入力を $\mathbf{x} \in \mathbb{R}^m$ 、出力を $\mathbf{y} \in \mathbb{R}^n$ とおく。効率のよい情報処理を行うためには、限られた n において \mathbf{y} ができるだけ多様な情報を表現できることが望ましい。これは、統計においては \mathbf{y} のエントロピー $H(\mathbf{y})$ を最大化することに相当する。ニューロン応答は確率的であるため、同一の入力に対しても揺らぎがある。これは \mathbf{x} が与えられた条件下においても \mathbf{y} がなおエントロピー $H(\mathbf{y}|\mathbf{x})$ を持つことに相当する。ただ、もしこの系が \mathbf{x} を用いて何らかの情報処理を行っているとするれば、この揺らぎは \mathbf{x} が変化したことによる変化よりは小さいと考えられる。 \mathbf{x} と \mathbf{y} の相互情報量は、

$$I(\mathbf{x}; \mathbf{y}) = H(\mathbf{y}) - H(\mathbf{y}|\mathbf{x}) \quad (1.5)$$

で与えられる。これは \mathbf{x} が与えられたことによる \mathbf{y} のエントロピーの減少量を表している。 $I(\mathbf{x}; \mathbf{y}) = 0$ であれば、 \mathbf{x} が与えられても \mathbf{y} のエントロピーは全く減少しない、つまり \mathbf{x} の情報は \mathbf{y} に全く反映されていないことになる。逆に $I(\mathbf{x}; \mathbf{y})$ が大きければ、多くの情報が伝達され、 \mathbf{y} のエントロピーが大きく減少したことになる。条件付きエントロピー $H(\mathbf{y}|\mathbf{x})$ (出力のノイズ) を減らすことができないとすると、入出力間の相互情報量 $I(\mathbf{x}; \mathbf{y})$ の最大化と $H(\mathbf{y})$ の最大化は等価になる。より直感的に言えば、入力に関わりのないノイズを増やすことなく出力を多様にすることは、入出力の情報伝達効率を上げることに等しい。

出力全体のエントロピーと個々の成分のエントロピーの間には、次の関係が成り立つ。

$$H(\mathbf{y}) \leq \sum_{i=1}^n H(y_i) \quad (1.6)$$

等号が成り立つのは \mathbf{y} の各成分が独立のときに限られる。右辺の各項は、各出力の多様性を表しているが、これらを大きくするには限界がある。例えば y_i をニューロンの発火頻度とすると、最大発火頻度を大きくすることで、 y_i のとり得る範囲を広げることになる。しかしこれは、ニューロンのエネルギー消費を増大することにつながる。従つて、個々の出力の多様性を固定するのであれば、全体の多様性を最大化するためには、各出力が独立であることが望ましい。

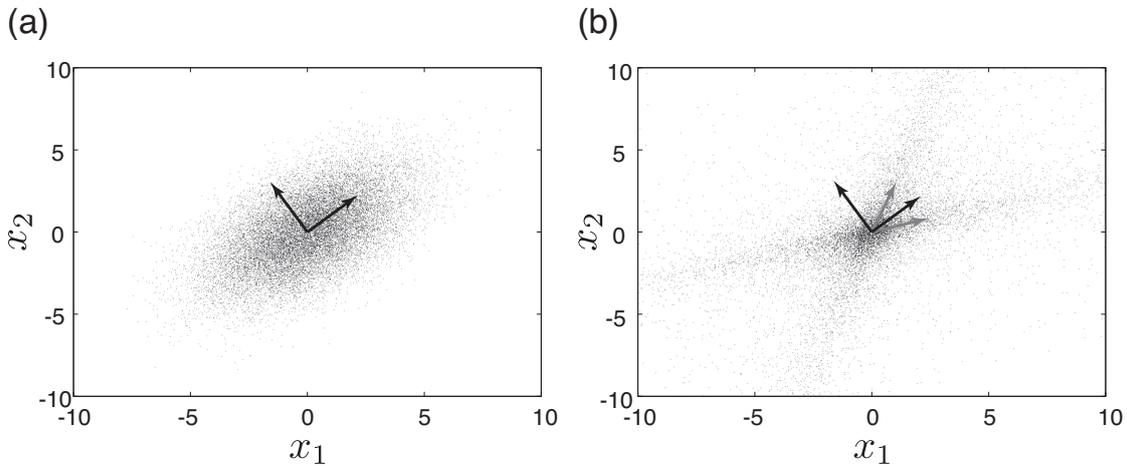


図 1.5: 情報量最大化原理による教師なし学習の例. (a) 入力分布が正規分布の場合. (b) 入力分布が非正規性を持つ場合. 黒の矢印は主成分分析 (PCA) で抽出できる直交基底. 灰色の矢印は PCA では抽出できない真の独立成分.

このような、限られた資源の下で表現する情報を最大化するという原理は、大脳皮質が教師なしでどのように学習すればよいかという問題への手掛かりを与える. McCulloch-Pitts のモデルや発火頻度モデルのように、シナプス結合による複数の入力の伝達を大まかに線型演算と捉えると、外側膝状体からの入力 \mathbf{x} に対する視覚野ニューロン集団の応答 $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$ は行列 $W = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n) \in \mathbb{R}^{m \times n}$ を用いて

$$\mathbf{y} = W^T \mathbf{x} \quad (1.7)$$

と書ける. \mathbf{x} が網膜像をほぼ再現しているとする. W の各列ベクトル \mathbf{w}_i は y_i で表される視覚ニューロンの受容野に相当する. この定式化の下で出力のエントロピー $H(\mathbf{y})$ を最大化することを考える. ただし, W^T を単に大きくすれば $H(\mathbf{y})$ を大きくすることができてしまうので, $\|\mathbf{w}_i\| = 1$ などの W への制約を加えることとする. 図 1.5(a) のように \mathbf{x} が多変量正規分布に従っている場合, 分散が大きい方向から順に W の列としていくことで, $H(\mathbf{y})$ を最大にできる. これは \mathbf{x} の分散共分散行列の固有値問題に帰着し, 数学や統計学において主成分分析 (principal component analysis; PCA), または分野によっては Karhunen-Loeve 展開と呼ばれる手法と等価である.

一方, 図 1.5(b) は正規分布より裾の重い分布から独立に生成した \mathbf{s} をある行列 V で変換したものであるが, この場合 \mathbf{y} は無相関であっても独立にならないから, $H(\mathbf{y})$ をより大きくする W^T が存在する. これは, $H(\mathbf{y})$ を最大にするという目的においては, PCA は必ずしも十分な手法ではないことを表している. これを解決

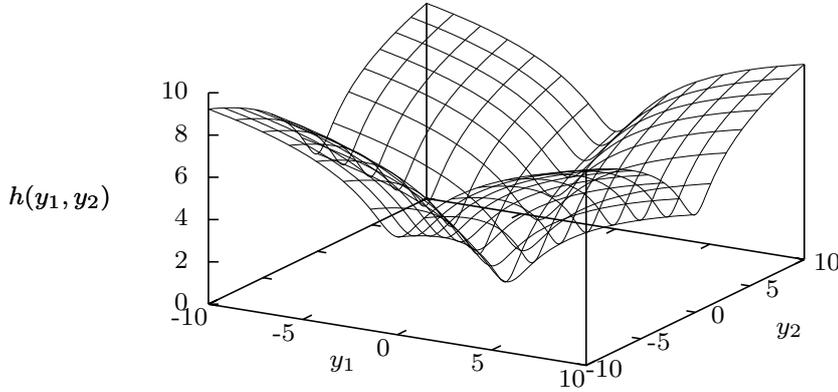


図 1.6: スパースネスの制約を与える関数 $h(\mathbf{y})$.

する手法として独立成分分析 (independent component analysis; ICA) [44] があり、後述するように PCA とともに脳の教師なし学習のモデルとしても重要である。ICA は、出力 \mathbf{y} の各成分のうち正規分布に従うものが高々一つである場合 (二つ以上の場合、上で述べた理由によりそれらの無相関性までしか保証できない) に、これらが独立になるような線型変換 W^T を見出す手法である。

ICA と深い関わりを持つ V1 の学習モデルに、Olshausen らのスパースコーディングがある [45, 46, 47]。通常の視覚経験の下で、視覚皮質のニューロンは常に活動しているわけではない。多くのニューロンは自発発火頻度に近い、低い発火頻度を保っており、視覚刺激に応じて少数のニューロンが発火している。つまり、視覚刺激に対する視覚野のニューロン集団の応答は「スパース」であるといえる。彼らはこの特性を再現するため、以下の目的関数を最小化するというモデルを提案した。

$$E = \frac{1}{2} \|\mathbf{x} - A\mathbf{y}\|^2 + \lambda h(\mathbf{y}) \quad (1.8)$$

第一項は、入力 \mathbf{x} と出力から再構成された入力の推定値 $A\mathbf{y}$ の差を表す。 $\lambda > 0$ で重み付けされた第二項 $h(\mathbf{y})$ は、 \mathbf{y} のスパースネスに関する制約項、すなわち、 \mathbf{y} がスパースでなければ大きくなるような関数である。このような関数の一つとして、Olshausen らは

$$h(\mathbf{y}) = \sum_j \log(1 + y_j^2) \quad (1.9)$$

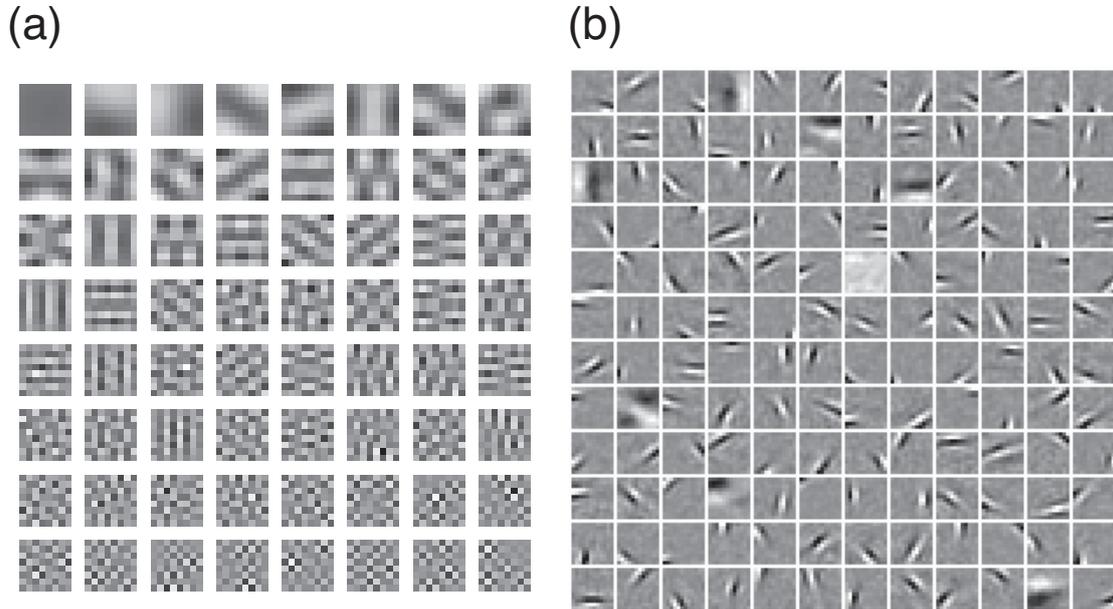


図 1.7: 自然画像の主成分分析 (a) とスパースコーディング (b). [45] より転載.

を用いている. 図 1.6 は \mathbf{y} が二成分の場合の $h(\mathbf{y})$ を示している. このグラフから, $h(\mathbf{y})$ が軸付近, すなわち y_1, y_2 のうち一方が 0 に近い領域で特に小さくなっていることがわかる. 従って, $h(\mathbf{y})$ を最小化することで, \mathbf{y} がスパースに分布するように制約することができる. これに対して, $\|\mathbf{y}\|^2$ のような \mathbf{y} のノルムの関数は, 超球面 $\{\mathbf{y} \mid \|\mathbf{y}\| = r\}$ 上で一定の値をとるため, $h(\mathbf{y}) = \|\mathbf{y}\|^2$ とした場合, \mathbf{y} はスパースにはならない.

モデルのパラメータである A と \mathbf{y} は, E の勾配に基づいて修正される. 出力の変化量 $\Delta\mathbf{y}$ は

$$\Delta\mathbf{y} \propto -\frac{\partial E}{\partial \mathbf{y}} = A^T \mathbf{x} - A^T A \mathbf{y} + \lambda \frac{\partial E}{\partial \mathbf{y}} h(\mathbf{y}) \quad (1.10)$$

となる. 第一項は入力 \mathbf{x} がシナプス結合 A^T を介して皮質ニューロンに線型に伝わることを表しており, 第二項はニューロン間の側方抑制を表している. また A の変化量 ΔA は

$$\Delta A \propto -\frac{\partial E}{\partial A} = \mathbf{x}\mathbf{y}^T - A\mathbf{y}\mathbf{y}^T \quad (1.11)$$

のように与えられ, 入出力間の Hebb 則や側方結合における anti-Hebb 則が導き出されていることがわかる.

このような学習則の下で自然画像の学習を行うことで, 図 1.7(b) のような, 方位選択性とバンドパス特性を併せ持つ表現が線型変換 A の基底として獲得される.

これは V1 でみられる受容野に類似している. 図 1.7(a) のように, 同じ自然画像に主成分分析を適用した結果では, 画素間の相関に基づく直交基底が抽出されるが, これらは局在化しておらず, V1 でみられる受容野とは大きく異なっている. このことからスパースコーディングは, V1 の自己組織的な受容野の獲得を説明するモデルであるといえる.

1.3.4 生成モデル

スパースコーディングをベイズ的な観点でみることもできる [48]. 式 (1.8) の第一項は, 誤差 $\mathbf{x} - A\mathbf{y}$ の成分毎の二乗和なので, 誤差の各成分が独立に正規分布に従っているとして, 対数尤度の符号を反転させたものに等しい. また式 (1.9) の各項は, 標準コーシー分布の確率密度関数

$$p(y_j) = \frac{1}{\pi(1 + y_j^2)} \quad (1.12)$$

の対数から定数項を除き, 符号を反転させたものに一致する. 従って, 式 (1.8) の最小化は, 誤差の分布 $p(\mathbf{x}|A, \mathbf{y})$ の正規性を仮定し, \mathbf{y} の事前分布 $p(\mathbf{y})$ として独立なコーシー分布を仮定した上で A, \mathbf{y} の事後確率

$$p(A, \mathbf{y}|\mathbf{x}) = p(\mathbf{x}|A, \mathbf{y})p(A)p(\mathbf{y}) \quad (1.13)$$

の最大化, すなわち MAP (Maximum a posteriori) 推定を行っていることに相当する. コーシー分布は正規分布に比べて非常に裾の重い分布である. このため, 標準コーシー分布に従う y_j は, ほとんどの場合 0 付近をとり, まれに外れ値のような大きな値をとる. このことから, 式 (1.8) が y_j のスパース性を仮定しているといえる.

$p(\mathbf{x}|\boldsymbol{\theta})$ のように, 入力 \mathbf{x} を生成する過程をパラメータ $\boldsymbol{\theta}$ で記述するモデルは生成モデルと呼ばれる. スパースコーディングの場合, $A\mathbf{y}$ に正規分布に従うノイズが重畳されたものとして, 入力 \mathbf{x} の生成過程を記述していることになる. ニューロン集団の応答に相当する \mathbf{y} は, A の各列で表される受容野形状の組み合わせで画像を表現し, その表現がスパースになるようにシナプス結合 A を変化させる.

スパースコーディングと ICA が深い関わりを持つのは, 両者が共に出力 \mathbf{y} の非正規性を仮定しているためである. \mathbf{y} が正規分布に従うと仮定すれば, スパースコーディングと ICA は PCA と等価になる (ただし, 入出力の次元の違いなどは除く). \mathbf{y} の各成分の分布が左右対称であるとする, 正規分布と尖度の異なる分布を仮定することが重要である. ICA では各 y_i が正規分布より尖度が大きい分布 (このような分布を super-Gaussian と呼ぶ) と小さい分布 (同じく sub-Gaussian)

のどちらかに従うとして、原信号を推定する線型変換を計算する。スパースコーディングは、全ての成分がsuper-Gaussianに従うと仮定している点において、ICAの特殊な場合であるといえる。

1.3.5 階層モデル

前述の通り、視覚系の処理は大まかに二つの経路に分かれて階層的に行われる。各経路に沿って初期の領野から高次領野に進むと、ニューロンの受容野は徐々に大きくなり、その選択性は複雑になって行く。一方、皮質領野間、またV1-LGN間には遠心性の結合も存在する[49]。

川人ら[32, 50]は、視覚系が直面している逆問題が、この双方向結合によって高速に計算されているとする、視覚皮質の計算理論を提案している。視覚系は二次元の網膜像から外界の三次元構造を推定しなければならない。外界の状態 \mathbf{y} が網膜像 \mathbf{x} として投影される過程を光学 A とすると、視覚系は、その逆変換である逆光学 A^{-1} を高速に計算する必要がある。彼らのモデルでは、 \mathbf{y} から A によって生成された \mathbf{x} を初期視覚野が受け取り、高次視覚野が \mathbf{y} の推定値 $\hat{\mathbf{y}}$ を表現する。ここで求心性結合は近似逆光学 $A^\#$ に相当し、 $\hat{\mathbf{y}}$ を計算する。 $\hat{\mathbf{y}}$ は最初、 \mathbf{y} に一致しないので、 A に相当する遠心性結合で $\hat{\mathbf{y}}$ から元の画像の推定値 $I\hat{\mathbf{x}}$ を計算し、誤差 $\hat{\mathbf{x}} - \mathbf{x}$ を再び求心性結合で伝搬することで、 $\hat{\mathbf{y}}$ を \mathbf{y} に近づけて行く。

スパースコーディングは、川人らの計算理論の一部を成していると解釈することができる。光学を $A = A_1 \circ A_2 \circ \dots$ のように複数の段階に分けられるとすると、V1のスパースコーディングは、最終段階である線分の表現から二次元画像への変換 A_1 の推定に対応する。Raoら[51, 52]はこれに非線型性を加え、上位の層を追加した予測符号化モデルを提案した。このモデルでは以下の目的関数を最小化する。

$$E = \sum_{l=1}^L \frac{1}{2\sigma_l^2} \|\mathbf{y}^{(l-1)} - f(A^{(l)}\mathbf{y}^{(l)})\|, \quad \mathbf{y}^{(0)} = \mathbf{x} \quad (1.14)$$

ここで添字 (l) は階層を表し、 σ_l^2 は各階層での誤差分散を表す。彼らは $L = 2$ のモデルに自然画像を学習させ、獲得された $A^{(2)}$ の基底が線分の組み合わせを表現することや、トップダウンの信号による応答の文脈修飾などのV1でみられる特性が再現できることを示した。また、著者ら[53, 54]は予測符号化モデルをさらに多層化し、高次視覚野における複雑なパターンの表現が、類似の枠組みを積み重ねることで実現できる可能性を示唆した。以上のように生成モデルは、受容野を自己組織的に獲得するメカニズムや文脈依存型の応答特性など、V1の特性の多くを説明するモデルであるといえる。

また、階層ベイズに基づく様々なモデルも提案されている。特に Hyvärinen ら [55, 56, 57, 58] の Topographic ICA (TICA) が有名である。このモデルは二つの階層からなる。第一層ではスパースコーディングと同様に入力画像から線型基底を抽出する。第二層では、第一層の出力の分散を表現する。彼らは第二層からのトップダウンの信号により、第一層で位置が近いニューロンが相関を持つように拘束条件を与え、複雑型細胞の応答特性が自然画像から学習できることを示した。また、第一層において V1 でみられる retinotopy が保存された。同じく階層ベイズに基づくモデルに Karklin [59, 60] のモデルがある。彼らもまた第二層において第一層の分散を表現するモデルを構築したが、彼らは第二層の基底を第一層と同様に学習させ、より複雑で大域的な特徴が獲得されることを示している。

1.4 初期視覚ニューロン応答の測定手法

測定技術の進歩によって、近年では様々な手法を用いてニューロンの活動を測定することができる。ニューロンが電氣的な活動をしていることは古くから知られていたため、これを微小電極を用いて直接計測するというアプローチは、初期の研究から行われてきた。また、電場やそれによって誘導された磁場が生体の組織をある程度透過することを利用し、脳波 (electroencephalography; EEG) や脳磁図 (magnetoencephalography; MEG) といった非侵襲な手法も用いられる。さらに、ニューロンの活動により生じる代謝を利用してニューロンの活動を間接的に計測する核磁気共鳴画像法 (magnetic resonance imaging; MRI) やポジトロン断層法 (positron emission tomography; PET) などの手法も、近年では用いられるようになった。

頭蓋骨の内部に微小電極を挿入する方法は、極めて侵襲的である。そのためヒトに関する電気生理的な知見は、医療目的で頭蓋内電極が装着されているてんかん患者などの特別な場合に限って得られ、てんかん発作など疾患そのものに関するものを除けば非常に限られている (例えば [61])。非侵襲な手法は間接的な計測であるために、微小電極を用いる手法に比べて時空間解像度が著しく劣る。特に、EEG や MEG は測定された電磁場から信号源を特定することが容易でないため空間解像度が、MRI や PET は代謝が起こるまでに時間がかかるために時間解像度が落ちるといったトレードオフの関係がある。このためニューロンや局所回路のレベルでは、動物を用いた電気生理実験が重要である。

微小電極を用いる手法は、細胞外電極による計測と細胞内電極による計測の二つに大別される。前者は、主に生体内の計測対象のニューロン付近に微小電極を挿入し、その先端における電位を計測するものである。また、後者は対象となるニューロンを標本として取り出し、その細胞膜内に挿入した微小電極によってニューロン

内部の電位を直接計測するものである。細胞内電極を用いた手法によって、シナプスやイオンチャネルの働きなど、単一ニューロンの詳細な特性が明らかになった。

細胞外電極による計測では、電極の先端は細胞間の空間に挿入されるので、測定される電位は計測対象のニューロンだけでなく、近傍の複数のニューロンの電気的な活動の影響を含むことになる。このため、分析の前処理として、スパイク信号の波形や細胞間を占める空間の電気的特性などの事前知識をもとに、対象ニューロンのスパイク信号を抽出するスパイクソーティングが行われることが多い。この場合、細胞外電極で計測される信号は、時刻を t とすると

$$S(t) = \sum_{i=1}^N \delta(t - \tau_i) \quad (1.15)$$

のように、スパイクの発生した時刻 τ_1, τ_2, \dots という形に縮約されることになる。光刺激が一定であったとしても、個々のスパイクの間隔は一定にはならない。この一見するとランダムなスパイク列が何を表現しているのかについては、現在でも大きな問題である。Adrian と Zotterman[62] が、刺激の強度に応じてスパイクの頻度が増加するという現象を発見して以来、ニューロンがスパイクの頻度によって情報を表現、または他のニューロンへ伝達しているという rate coding モデルが実験、理論を問わず多くの研究の基礎となっている。一方でスパイク列の時間的構造が情報を表現しているという temporal coding も提案されており、脳の部位によっては実際に確認されている（例えば [63]）。どちらの場合においても、このようにランダム性を持つデータから有用な情報を抽出するためには、何らかの統計的な処理を用いることが不可欠である。

1.4.1 逆相関法

逆相関法は、ランダムパターンを次々と呈示することでニューロンを刺激し、パターン列とスパイク列との相関を取ることで、非線形な応答特性に隠された受容野を測定する手法である [64, 65]。また、高次の相関を用いて非線形なカーネルを推定することも可能である [66]。

後述するように、呈示刺激 $\mathbf{x} \in \mathbb{R}^m$ に対するニューロンの発火頻度 $R(\mathbf{x})$ が線型フィルタ $\mathbf{w} \in \mathbb{R}^m$ と非線型関数 $N: \mathbb{R} \rightarrow \mathbb{R}_+$ を用いて $R(\mathbf{x}) = N(\mathbf{w} \cdot \mathbf{x})$ で近似できるとき、スパイクを生起させた刺激の総和をとることで \mathbf{w} を推定することができる。ここで、関数 N は測定対象のニューロンにおけるスパイク生成過程、 N への入力 $g = \mathbf{w} \cdot \mathbf{x}$ はシナプス入力総和と考えられる。一方、 g の生成過程は一般には非線型である。そこで g の生成過程を $g = F(\mathbf{x})$ とし、この関数の多項式展開

$$F(\mathbf{x}) = F_0 + \sum_i F_{1,i} x_i + \sum_{i,j} F_{2,ij} x_i x_j + \dots \quad (1.16)$$

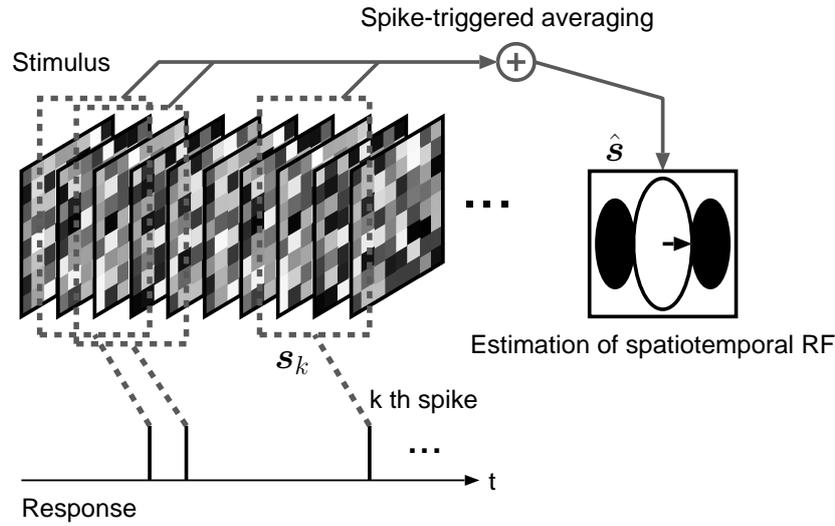


図 1.8: 逆相関法によるニューロンの時空間受容野の測定. RF: receptive field.

を考える. $F_{n,ij\dots}$ を n 次カーネルと呼ぶ. 上で述べた線型逆相関法は一次カーネルの推定に相当する. また近年では, 二次以上の非線形特性を分析するための様々な手法が提案されている. たとえば, スパイクを生成させた刺激の共分散 (STC; spike-triggered covariance) を用いた分析手法 [66] などがある.

逆相関法は視覚系の電気生理実験以外でも用いられている. 例えば聴覚系における生理実験 [67] や心理実験 [68, 69] における応用例もある.

初期の逆相関法では呈示刺激としてホワイトノイズが用いられていた. しかし高次のニューロンではホワイトノイズで十分に頻度の高い応答が得られないという問題がある. そこで近年では, 呈示刺激に自然画像を用いた分析も提案されている [70, 71]. また, 自然画像から抽出した基底を利用して呈示刺激を生成する方法も提案されている [72].

1.4.2 逆相関法の理論

逆相関法では得られたデータに対して以下の計算を行う [66].

$$\hat{\mathbf{x}} = \frac{1}{K} \sum_{k=1}^K \mathbf{x}_k \quad (1.17)$$

この手順を図 1.8 に示す. 灰色の破線で示すように, ベクトル \mathbf{x}_k は k 番目のスパイクに先行する一定期間に呈示された刺激を表す. \mathbf{x}_k の要素は各フレーム, 各ピクセ

ルの輝度値など、視覚刺激の成分を表している。 $\hat{\boldsymbol{x}}$ を STA (spike-triggered average) と呼ぶ。十分なスパイク数 K が得られたとき、STA の計算結果として時空間受容野が得られる。

Chichilnisky[65] は、逆相関法により受容野が測定できる原理と、その条件を示している。以下ではその概要を示す。まず、刺激 \boldsymbol{x} は球対称 ($\forall \boldsymbol{x}, \boldsymbol{x}^*, |\boldsymbol{x}| = |\boldsymbol{x}^*| \Rightarrow P(\boldsymbol{x}) = P(\boldsymbol{x}^*)$) に分布する必要がある。ホワイトガウスノイズはこの条件を満たしている。また、ニューロンの応答特性 $R(\boldsymbol{x})$ を刺激 \boldsymbol{x} に対する発火頻度で定義すると、 $R(\boldsymbol{x})$ は次の式で表せるとする。

$$R(\boldsymbol{x}) = N(\boldsymbol{w} \cdot \boldsymbol{x}) \quad (1.18)$$

ここで、 \boldsymbol{w} は \boldsymbol{x} と同じ次元のベクトルである。式 (1.18) は、ニューロンが \boldsymbol{x} の各要素の線形和 $g = \boldsymbol{w} \cdot \boldsymbol{x}$ (g を generator signal と呼ぶ) と一変数関数 $N(g)$ によって応答を決定していることを仮定している。 \boldsymbol{w} の各要素は、過去のある時刻にある位置に呈示された光の刺激が、現在の発火頻度にどのように影響するかを表している。この \boldsymbol{w} を時空間受容野と呼ぶ。 $N(g)$ は任意の関数で、一般には非線形である。この $N(g)$ を static nonlinearity と呼ぶ。このように、入力値の線形和と非線形関数の組み合わせで構成されるニューロン応答のモデルを LN (linear-nonlinear) モデルと呼ぶ。以上の条件を満たす場合、十分に長い期間の計測を行うと、STA は

$$\hat{\boldsymbol{x}} \propto \boldsymbol{w} \quad (1.19)$$

となり、 $N(g)$ の形状によらず時空間受容野 \boldsymbol{w} に相似なベクトルが得られる。

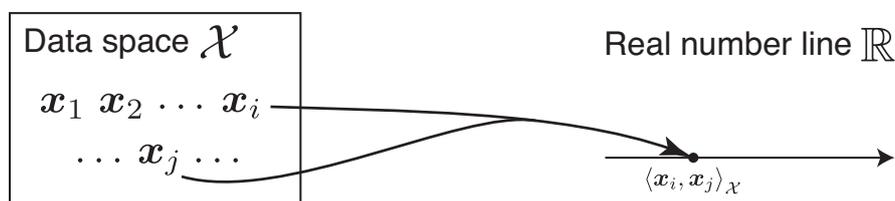
また、既存の実験データを用いて static nonlinearity の推定ができる [73]。刺激 \boldsymbol{x} に対する generator signal は、得られた受容野 \boldsymbol{w} と \boldsymbol{x} の内積をとることで得られる。応答 $N(g)$ は既に得られているので、 g を求めることによって $N(g)$ が推定できたことになる。

1.5 カーネル法

多変量解析では、データが属する空間における内積や距離によって計算手法が与えられることが多い。例えば 1.3.4 節で述べたスパースコーディングのように、推定誤差を最小化するモデルでは、入力ベクトルが属する空間で定義された距離に基づいて目的関数が定義される。カーネル法は、データを内積や距離が定義された高次元空間に写像することで、より複雑な解析を同様の計算手法で行うことを可能にする枠組みである [74]。

図 1.9 にカーネル法の概念を示す。データの空間を \mathcal{X} とすると、 \mathcal{X} の元 $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots$ をいくつかのクラスタに分離、またはこれらを代表する点を求める場合、解の善

(a)



(b)

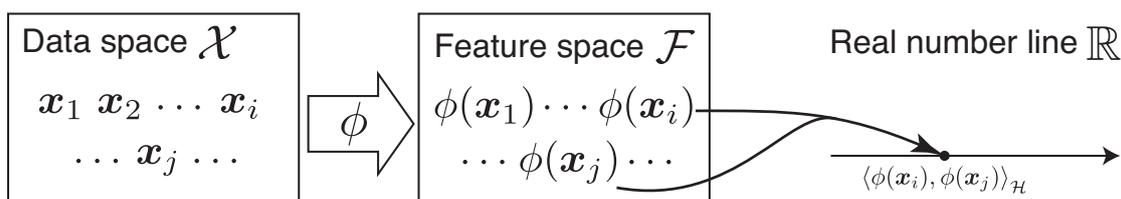


図 1.9: カーネル法の概念図. (a) データ空間 \mathcal{X} の計量に基づく手法. (b) カーネル法. 写像 ϕ によって特徴空間に移されたデータの集合に対して (a) を適用する.

し悪しを決める基準として、二点間の距離が重要である. 代表点の例では、代表点 $\mathbf{p} \in \mathcal{X}$ と $\mathbf{x}_1, \mathbf{x}_2, \dots$ との距離が平均して小さくなるように \mathbf{p} が決定される. 通常の距離空間 \mathcal{D} では、距離 $d(\mathbf{x}, \mathbf{y})$ は

$$d_{\mathcal{D}}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_{\mathcal{D}} = \langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{D}} \quad (1.20)$$

のようにノルムと内積で表される. 従って、データの空間に内積が定義されていれば、任意の二点間の距離を定義することができる、もし既存のアルゴリズムが距離やノルム、内積の概念のみを用いて構成されていれば、全く同等の手法を別の距離空間に属するデータに適用することができる (図 1.9(a)).

カーネル法では、この内積 $\langle \cdot, \cdot \rangle$ の代わりにカーネル関数と呼ばれる二変数の関数 $k(\cdot, \cdot)$ を用いる [75, 76]. 後述するが、 $k(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ が一定の条件を満たせば、 $k(\cdot, \cdot)$ に対して \mathcal{X} とは別の空間 \mathcal{F} と写像 $\phi : \mathcal{X} \rightarrow \mathcal{F}$ が存在して

$$k(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle_{\mathcal{F}} \quad (1.21)$$

とできる. 右辺は \mathbf{x} と \mathbf{y} を ϕ によって \mathcal{F} に写像し、 \mathcal{F} において内積を計算したことに等しい (図 1.9(b)). この \mathcal{F} を特徴空間と呼ぶ.

カーネル法が有用であるのは、特徴空間の次元が元の空間よりも大きくなることにより、データの表現能力が向上するためである. 例えばサポートベクターマシン (SVM) [77] は、図 1.10 のようにデータを二つのクラスに線型分離する学習

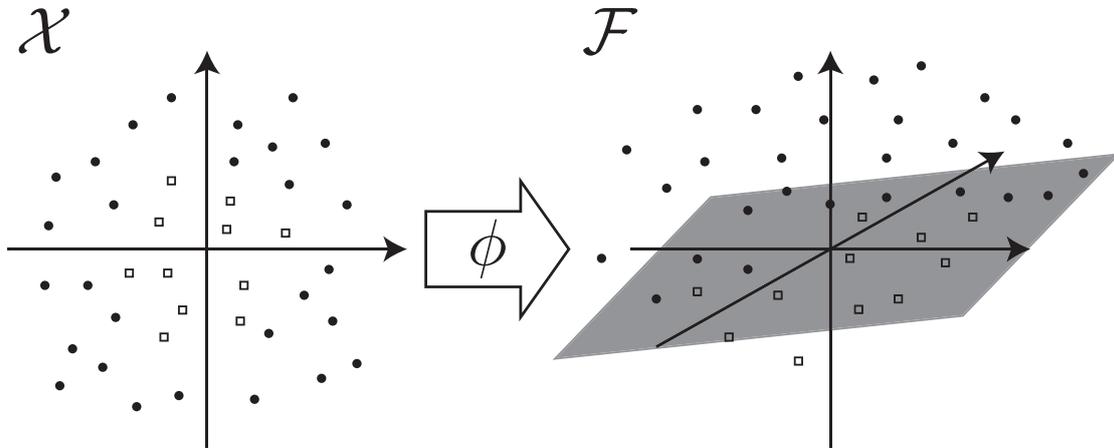


図 1.10: サポートベクターマシンにおけるカーネル法の適用.

機械であるが，データが線型分離不可能である場合は，そのままでは用いることができない．しかし，データを十分大きい次元を持つ特徴空間 \mathcal{F} に写像することで，複雑な境界を持ったデータを \mathcal{F} 上で分離する超平面を見つけることができる．

特徴空間 \mathcal{F} の次元が大きい場合， \mathcal{F} への写像と \mathcal{F} 上での内積の計算は長い計算時間を要する．カーネル法では，これらの合成と等価なカーネル関数を短時間で計算することで，表現力の高い特徴空間における演算を効率的に行うことができる．これはカーネルトリックと呼ばれている．

1.5.1 カーネル関数の条件

カーネル関数が以上のような性質を持つためには，次のような条件がある．まず，内積の対称性より明らかに

$$k(\mathbf{x}, \mathbf{y}) = k(\mathbf{y}, \mathbf{x}) \quad (1.22)$$

でなければならない．有限個のベクトルの集合 $\{\mathbf{x}_k\}_{k=1}^n$ から任意の二つのベクトルを選び， $\mathbf{x}_i, \mathbf{x}_j$ とする．全ての i, j の組み合わせについて $k(\mathbf{x}_i, \mathbf{x}_j)$ を求め， $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ のように配置した行列はグラム行列と呼ばれる．カーネル関数を特徴ベクトルの内積としてみると， $K_{ij} = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle_{\mathcal{H}}$ と書ける．ここで任意の n 次元ベクトル \mathbf{c} に対する二次形式 $\sum_{i,j} K_{ij} c_i c_j$ を考えると，内積の双線形性と正定値

性より

$$\begin{aligned} \sum_{i,j} K_{ij} c_i c_j &= \left\langle \sum_{i=1}^n c_i \phi(\mathbf{x}_i), \sum_{j=1}^n c_j \phi(\mathbf{x}_j) \right\rangle_{\mathcal{H}} \\ &= \left\| \sum_{i=1}^n c_i \phi(\mathbf{x}_i) \right\|_{\mathcal{H}}^2 \geq 0 \end{aligned} \quad (1.23)$$

が常に成り立たなければならない。逆に、式(1.22)と式(1.23)を満たす \mathcal{X} 上の二変数関数 $k: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ は、対応する特徴空間が存在する。

1.5.2 カーネル関数の構成

式(1.23)は、任意の $n \in \mathbb{N}$ に対して n 個のベクトルをどのように選んでも、そのグラム行列が半正定値対称行列となることを意味している。これは、同じ二変数関数である内積の正定値性 $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ よりも厳しい条件となっている。

そこで、既存のカーネル関数を組み合わせて新たなカーネル関数を構成することが行われる。いま、二つの正定値カーネル $k^{(1)}, k^{(2)}$ (導関数ではない)があり、対応する特徴空間への写像が $\phi^{(1)}, \phi^{(2)}$ であるとして、これら二つのカーネル関数から新たなカーネル関数を構成する方法をいくつか述べる。

二つのカーネル関数の和

$$k(\mathbf{x}, \mathbf{y}) = k^{(1)}(\mathbf{x}, \mathbf{y}) + k^{(2)}(\mathbf{x}, \mathbf{y}) \quad (1.24)$$

は正定値カーネルとなり、対応する特徴空間への写像は

$$\phi(\mathbf{x}) = \begin{pmatrix} \phi^{(1)}(\mathbf{x}) \\ \phi^{(2)}(\mathbf{x}) \end{pmatrix} \quad (1.25)$$

と表せる。 $\phi^{(1)}, \phi^{(2)}$ が有限次元ベクトルの場合、 ϕ の次元はこれらの次元の和となる。また、積

$$k(\mathbf{x}, \mathbf{y}) = k^{(1)}(\mathbf{x}, \mathbf{y}) k^{(2)}(\mathbf{x}, \mathbf{y}) \quad (1.26)$$

もまた正定値カーネルとなり、特徴空間は $\phi^{(1)}$ と $\phi^{(2)}$ のテンソル積になる。

1.5.3 カーネル関数の例

カーネル関数の簡単な例として、多項式カーネル

$$k^{\text{poly}}(\mathbf{x}, \mathbf{y}) = (c + \mathbf{x}^T \mathbf{y})^d \quad (1.27)$$

を考える. ここで $c \geq 0, d \in \mathbb{N}$ である. カーネル関数として意味を持つ最も単純な例を考えるために, $d = 2$ とする. $\mathcal{X} = \mathbb{R}^2$ とすると, $k^{\text{poly}}(\mathbf{x}, \mathbf{y})$ は

$$k^{\text{poly}}(\mathbf{x}, \mathbf{y}) = c^2 + 2cx_1y_1 + 2cx_2y_2 + x_1^2y_1^2 + 2x_1x_2y_1y_2 + x_2^2y_2^2 \quad (1.28)$$

のように展開できる. 式 (1.28) の各項を $\phi_i^{\text{poly}}(\mathbf{x})\phi_i^{\text{poly}}(\mathbf{y})$ の形に分解すると,

$$\phi^{\text{poly}}(\mathbf{x}) = (c, \sqrt{2cx_1}, \sqrt{2cy_1}, x_1^2, \sqrt{2x_1x_2}, x_2^2)^T \quad (1.29)$$

が得られる. このとき $\phi^{\text{poly}}(\mathbf{x}), \phi^{\text{poly}}(\mathbf{y}) \in \mathbb{R}^6$ の通常の内積 $\phi^{\text{poly}}(\mathbf{x})^T \phi^{\text{poly}}(\mathbf{y})$ が式 (1.28) に一致することが確かめられる. 従って, $k^{\text{poly}}(\mathbf{x}, \mathbf{y})$ の場合, 2次元のデータに対して6次元の特徴ベクトルが対応することになる. また, 特徴ベクトルには \mathbf{x} の一次の成分だけでなく二次の成分も含まれており, 一次の成分と二次の成分の相対的な重みが c によって制御されることがわかる. d をより大きな整数にすることで, 特徴ベクトルはより高い次数の成分を持つ高次元のベクトルになる.

特徴ベクトル $\phi^{\text{poly}}(\mathbf{x})$ の各成分を \mathbf{x} の関数とみると,

$$(L^{\text{poly}} \mathbf{e}_i)(\mathbf{x}) = \langle \mathbf{e}_i, \phi(\mathbf{x}) \rangle_{\mathbb{R}^6} = \phi_i^{\text{poly}}(\mathbf{x}) \quad (1.30)$$

より, これらが L^{poly} の像空間を張る基底となる. 従って, 多項式カーネル $k^{\text{poly}}(\mathbf{x}, \mathbf{y})$ によって定義される再生核ヒルベルト空間は, \mathbb{R}^2 上の d 次の多項式関数全体の集合となる.

1.6 本研究の目的

1.2.2 節で述べた通り, 眼球から入力された視覚情報に対して, 腹側経路では物体の形状や色などの処理が, 背側経路では主に物体の位置や動きに関する処理が行われる. 各経路はさらに複数の領野に分けられ, これらが双方向的に連絡し合っ
て階層処理を行っている. 各領野のニューロンがどのような特徴に応答するかは, IT (例えば顔ニューロン [29]) に至るまで部分的にはわかってきている. 一方, このような特徴選択性がどのように実現されているかについての理解は限定的である. 例えば V1 の単純型細胞の方位選択性は on-center, off-center 型受容野の組み合わせによる Gabor 型受容野によって説明できるが, V4 における曲率の表現 [25] や IT における顔に対する選択性の実現に関するはつきりとした理解は得られていない.

初期視覚野のニューロン応答特性は, 比較的低次の非線型性を仮定することで, 逆相関法を用いて推定することができる. しかし, より高次の非線型応答特性を分析するための統一的な手法は存在しない. そこで第二章では, 本手法では逆相関法

を用いて区分線形関数の局所的な線形カーネルを推定する手法を提案する。また、初期視覚野ニューロンのモデルを用いた数値実験を行い、得られた結果がニューロン応答特性のどのような側面を反映しているのか検証する。

V1の方位選択性を持つニューロンは、空間位相に選択性を持つ単純型細胞と、これを持たない複雑型細胞の二種類に分けることができる。これまでの様々な生理学的実験によって、これらのニューロンの詳細な応答特性に関する知見が得られている。しかし、この機能がどのように獲得されているのかという疑問を解決することも必要である。特に本研究では、複雑型細胞-単純型細胞間の偶対称な結合がどのように獲得されているのかに直目する。第三章では、階層ベイズモデルの第一層に非負性を導入することにより、単純型細胞応答を模したモデルを構築し、このような結合を自然画像の統計的性質から獲得できる可能性について検討する。

初期視覚領野における受容野の自己組織的獲得や、領野間の双方向性結合の役割は、生成モデルの枠組みで説明されることが多い。特に第一次視覚野が表現している線型基底の自己組織的獲得については、生成モデルでよく再現することができる。一方で、視覚野のニューロンはexpansiveな非線型性を持つことが知られている [2]。従ってこの非線型性の役割についても理解する必要がある。第四章では非負制約のある生成モデルにカーネル法を導入し、ニューロン応答の非線型性が応答特性や受容野獲得に及ぼす影響について検証する。

第2章 初期視覚野の非線型ニューロン応答の分析手法

2.1 目的

1.4節で述べた通り、初期視覚野における電気生理実験では、対象の感覚器に刺激を与え、微小電極を挿入してニューロンの活動電位（スパイク）を測定することが行われる。このとき、ニューロンは入力刺激に応じて応答を変化させるので、その入出力の関係を調べるのが実験の目的となる。

初期視覚野のニューロン応答特性は、線形フィルタまたは二次の非線形フィルタとして説明される。例えば複雑型細胞は、二次の非線形性を持ったエネルギーモデルで説明される。しかし、実際の複雑型細胞では、エネルギーモデルでは説明できない結果も得られる [78]。したがって、初期視覚野の応答特性は、より複雑なニューロン間の相互作用によって実現されていると考えられる。そこで、ニューロンの応答特性をより詳細に分析することで、その応答特性がどのように実現されているのかが調べられると期待できる。しかし、高次の非線形カーネルを逆相関法によって推定するには、多くの実験データと計算が必要になる。さらに、高次のカーネルが推定できても、それらの持つ意味を解釈することは難しい。

本章では、ニューロンの応答特性が区分線形で近似できると仮定し、高い非線形性を持つニューロン応答特性を分析する手法について検討する [79]。区分線形関数による関数の近似に関しては、Gauss-Newton法を利用した手法も提案されている [80]。しかし、ニューロンを対象にする場合、入力の次元が非常に大きいことと、得られる応答が確率的であることから、このような手法を直ちに適用することは難しい。そこで、本手法では逆相関法を用いて区分線形関数の局所的な線形カーネルを推定する手法を提案する。また、本手法では初期視覚野ニューロンのモデルを用いた数値実験を行い、得られた結果がニューロン応答特性のどのような側面を反映しているのか検証する。

2.2 ニューロン応答の区分線形近似

ニューロンの非線形性の例として、初期視覚野で見られる単純型細胞について考える。単純型細胞は特定の方位と空間周波数に選択的に応答する細胞である。最適刺激が呈示されたとき、単純型細胞の発火頻度はそのコントラストに応じて高くなる。また、無刺激の状態での自発発火率は低く、最適刺激を反転させた刺激には応答しない。この特性は、特定の方向に延びた受容野形状 \mathbf{w} と static nonlinearity $N(g)$ からなる LN モデルで説明される。Anzai ら [2] は STA によって推定した \mathbf{w} から各入力に対する $g = \mathbf{w} \cdot \mathbf{x}$ を求めて単純型細胞の static nonlinearity の形状を測定し、得られた結果を次式のような half-power 関数を用いて説明している。

$$N(g) = [g]_+^n \quad (2.1)$$

$$= \max(0, g^n) \quad (2.2)$$

実際は、発火頻度の上昇が始まる g の閾値や全体の発火頻度の大きさを表すため、さらに二つのパラメータが用いられているが、ここでは簡単のため省略する。

単純型細胞の受容野を \mathbf{w} とすると、応答特性は

$$R(\mathbf{x}) = [\mathbf{w} \cdot \mathbf{x}]_+^n \quad (2.3)$$

と表せる。この単一ニューロンは LN モデルで表せるため、逆相関法を適用すると、 n の値に関わらず \mathbf{w} に相似な STA が得られる。しかし、複数のニューロンを組み合わせたネットワークを考えると、generator signal の生成過程 $F(\mathbf{x})$ に他のニューロンの static nonlinearity が内在していることが確認できる。最も単純な例として、3 個のニューロンからなるフィードフォワードのネットワークを考える。上位ニューロンの特性を次式のように表す。

$$F(\mathbf{x}) = a_1[\mathbf{w}_1 \cdot \mathbf{x}]_+^{n_1} + a_2[\mathbf{w}_2 \cdot \mathbf{x}]_+^{n_2} \quad (2.4)$$

添字は下位の二つのニューロンに対応する。このように、非線形な項が複数含まれる場合、LN モデルで表すことは不可能であり、各項の非線形性が STA の形状に影響することになる。 $[g]_+^n$ は原点に対して非対称な関数であるため、その多項式展開には様々な次数の項が現れることになる。しかし、刺激 \mathbf{x} の空間を適切に区分すると、 $F(\mathbf{x})$ は簡単な局所関数の組として表すことができる。

$n \approx 1$ の場合、単純型細胞の特性は以下のように多変量区分線形関数で近似できる。

$$f_1(\mathbf{x}) = [\mathbf{w} \cdot \mathbf{x}]_+ \quad (2.5)$$

この関数 $f_1(\mathbf{x})$ は図 2.1 左上に示すように、 $(M-1)$ 次元超平面 $\mathbf{w}_1 \cdot \mathbf{x} = 0$ で分けられた二つの領域でそれぞれ線形である。また、 \mathbf{w}_1 と異なる受容野 \mathbf{w}_2 を持つ別の単

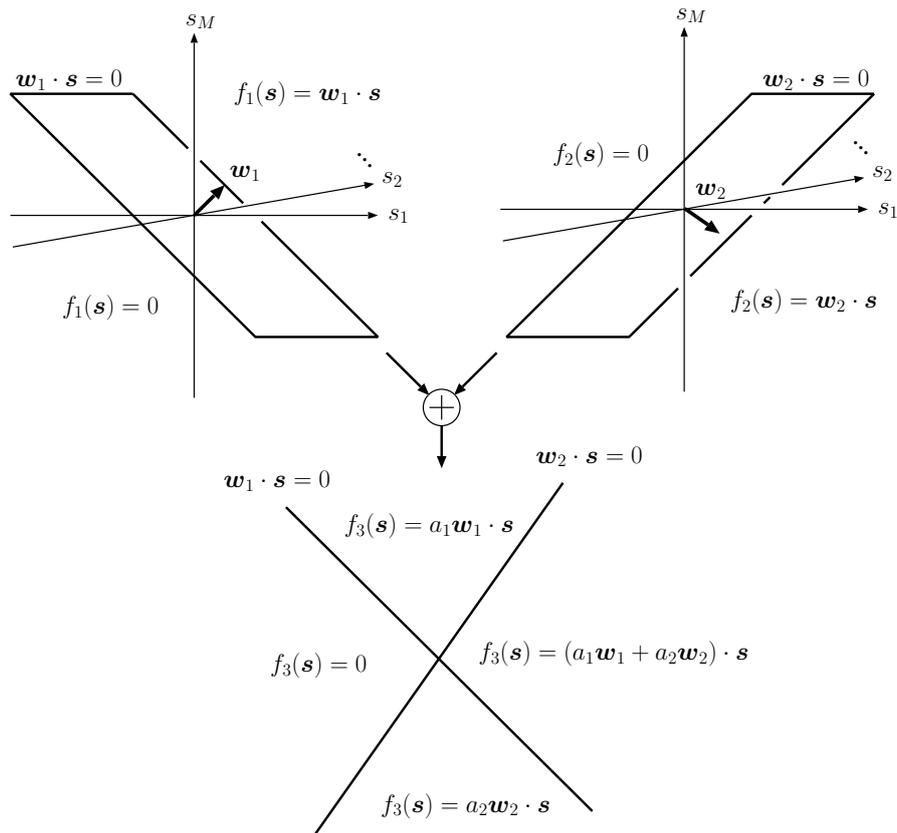


図 2.1: 多変量区分線形関数の例. 上図は二つの単純型細胞の特性 $f_1(\mathbf{x}), f_2(\mathbf{x})$ を表す. 下図はこれらの線形和 $f_3(\mathbf{x}) = a_1 f_1(\mathbf{x}) + a_2 f_2(\mathbf{x})$ を表す. 下図では軸を省略している.

純型細胞を考えると、その特性 $f_2(\mathbf{x})$ は異なる境界 $\mathbf{w}_2 \cdot \mathbf{x} = 0$ をもつ区分線形関数で表せる。これを図2.1右上に示す。ここで、これら二つの単純型細胞から興奮性の入力を受けるニューロンを考える。このニューロンの特性 $f_3(\mathbf{x}) = a_1 f_1(\mathbf{x}) + a_2 f_2(\mathbf{x})$ は、図2.1右のように4つの領域に分けられる。 $f_3(\mathbf{x})$ は各領域内では線形で、そのカーネルは単純型細胞の受容野 $\mathbf{w}_1, \mathbf{w}_2$ の線形和になっている。また、 $n \neq 1$ の場合も、より細かく領域を区分することで区分線形近似することができ、各領域のカーネルは $\mathbf{w}_1, \mathbf{w}_2$ の線形和で表せる。

以上の議論から、局所線形カーネルを求めることで、ニューロン応答特性を構成する線形フィルタ特性に関する情報が得られることが期待できる。次節ではこの局所線形カーネルを推定する方法について考察する。

2.3 逆相関法による局所線形カーネルの推定

ニューロンの応答特性 $R(\mathbf{x})$ を、区分線形関数 $F(\mathbf{x})$ を用いて

$$R(\mathbf{x}) = N(F(\mathbf{x})) \quad (2.6)$$

と近似することを考える。また、 $\mathbf{x} \in D \subset \mathbb{R}^M$ とする。区分線形関数 $F(\mathbf{x})$ の例を図2.2に示す。定義域 D を複数の領域 Q_1, Q_2, \dots に区分すると、 $F(\mathbf{x})$ は各領域における局所線形関数の組み合わせで表すことができる。領域 Q_p における局所線形関数を

$$F_p(\mathbf{x}) = \mathbf{w}_p \cdot \mathbf{x} + v_p \quad (2.7)$$

とする。ここで、 v_p は定数項である。 \mathbf{w}_p は Q_p における \mathbf{x} の各要素の重みを表している。この \mathbf{w}_p を局所線形カーネルと呼ぶ。また、 \mathbf{w}_p は Q_p における $F(\mathbf{x})$ の勾配 $\nabla F(\mathbf{x})$ を表している。

いま、図2.2のように、領域 Q_p 内に分布するパターンをランダムに生成し、ニューロンを刺激することを考える。刺激 \mathbf{x} の分布は、平均 \mathbf{m} を中心に球対称な分布とする。 $\mathbf{x} \in Q_p$ のとき、 $F(\mathbf{x}) = F_p(\mathbf{x})$ となる。このとき、 $N_p(g) = N(g + v_p)$ とおくと、

$$R(\mathbf{x}) = N(\mathbf{w}_p \cdot \mathbf{x} + v_p) = N_p(\mathbf{w}_p \cdot \mathbf{x}) \quad (2.8)$$

となる。この式は、式(1.18)と同様にLNモデルを表している。したがって、 \mathbf{m} に重畳した球対称なノイズのSTAを求めることで局所線形カーネル \mathbf{w}_p を推定することができる。スパイク k に先行する刺激を \mathbf{x}_k とすると、局所線形カーネルの推定値は次式のように求められる。

$$\hat{\mathbf{x}} = \frac{1}{K} \sum_{k=1}^K (\mathbf{x}_k - \mathbf{m}) = \frac{1}{K} \sum_{k=1}^K \mathbf{x}_k - \mathbf{m} \quad (2.9)$$

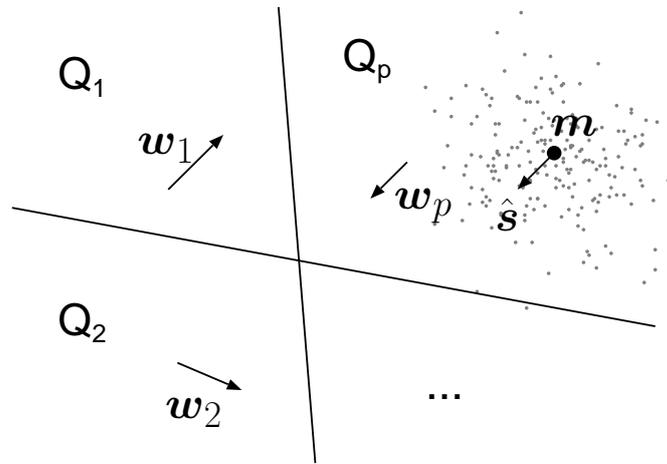


図 2.2: 局所線形カーネルの推定.

M 次元正規分布は球対称の条件を満たす確率分布の一つである。そこで、 \mathbf{x} の分布に平均 \mathbf{m} 、分散共分散行列 $\sigma^2 I$ の M 次元正規分布を用いる。 I は M 次の単位行列である。 σ^2 が十分小さければ、ほぼ全ての刺激は一つの領域に分布することになる。

ニューロンの応答特性が滑らかな関数 $f(\mathbf{x})$ である場合、小さな領域に区分された $F(\mathbf{x})$ によって $f(\mathbf{x})$ を近似することになる。ある \mathbf{x}_p を中心に領域 Q_p を小さくしていくと、 w_p は \mathbf{x}_p における $f(\mathbf{x})$ の勾配 $\nabla f(\mathbf{x}_p)$ に近づいていく。したがって、 \mathbf{m} に重畳するノイズの分散 σ^2 を小さく設定することで、 \mathbf{m} の周りの勾配を正確に推定できる。ただし、 σ^2 が極端に小さいとき、呈示刺激は \mathbf{m} からほとんど変化しないため、実用的な測定時間では推定値が収束しない。

対象のニューロンに関して事前知識がある場合は、 \mathbf{m} や σ^2 の値を決定することができる。しかし、一般に実験者は対象のニューロンの応答特性についての情報を持たない。そこで、以降の数値実験では \mathbf{m} にホワイトノイズを用い、複数の局所線形カーネルを求めることで、ニューロン応答特性の分析を行う。また、 σ^2 は \mathbf{m} の分散より十分小さい値を選択する。

2.4 数値実験

この節では、これまで述べてきた方法によって、実際に局所線形カーネルが推定できることを示すために数値実験を行う。実験対象には単純型細胞や複雑型細胞の応答特性を模したモデルを用いる。また、得られた局所線形カーネルがモデルの応答特性のどのような側面を反映しているかについて考察する。

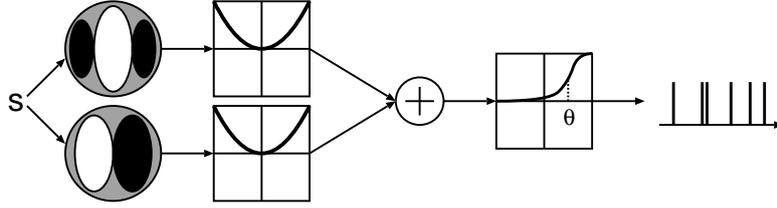


図 2.3: 複雑型細胞のモデル.

2.4.1 複雑型細胞モデル

複雑型細胞は単純型細胞と同様に、特定の時空間周波数に選択性を持つ細胞である。しかし単純型細胞と異なり、位相には選択性を持たない。この特性（ここでは空間特性のみを考える）は、以下のように二つの線形フィルタの組み合わせで説明される。

$$R(\mathbf{x}) = \sqrt{(\mathbf{w}_{\omega,0} \cdot \mathbf{x})^2 + (\mathbf{w}_{\omega,\frac{\pi}{2}} \cdot \mathbf{x})^2} \quad (2.10)$$

このモデルはエネルギーモデルと呼ばれる。 $\mathbf{w}_{\omega,\phi}$ は空間周波数 $\omega = (\omega_x, \omega_y)$ と空間位相 ϕ に選択性を持つ線形フィルタを表し、各要素には二次元の正弦波パターンにガウス窓をかけたガボール関数が用いられる。式 (2.10) は局所フーリエ変換における空間周波数 ω のパワーに対応する。そのため、複雑型細胞の位相に対して頑健な特性をよく再現する。しかし、実際の初期視覚野においてこの計算がどのように実現されているかについては、より詳しい研究が必要である。2.2 節で述べた単純型細胞の応答特性を考慮すると、より現実的には次のようなモデルを考えることができる。

$$R(\mathbf{x}) = \left[\sum_i a_i [\mathbf{w}_{\omega,\phi_i} \cdot \mathbf{x}]_+^{n_i} \right]_+^n \quad (2.11)$$

ここで、 a_i は各フィルタの重みである。 $i \in \{0, 1, 2, 3\}$, $\phi_i = \frac{\pi i}{2}$, $a_i = 1$, $n_i = 2$, $n = \frac{1}{2}$ の場合、式 (2.11) はエネルギーモデルと等価になる。各フィルタが単一の単純型細胞で実現されていると考えると、複雑型細胞の応答特性は 90° 間隔で位相が異なる 4 つの単純型細胞で実現できることになる。Anzai らが 39 個の単純型細胞について n_i を求めた結果、平均 2.17、標準偏差 0.53 で分布した。したがって、実際の単純型細胞も、平均的には二次の非線形性を持つといえる。しかし、この指数の推定結果には、ばらつきもみられる。このことから、複雑型細胞の応答特性は、機能的にはエネルギーモデルと同等になるように、より多くの単純型細胞を適度に組み合わせることによって実現されていると考えられる。

ここでは、本研究で提案している手法を上記のモデルに適用し、局所線形カーネルが得られることを確認するとともに、 n_i の値による結果の違いについて議

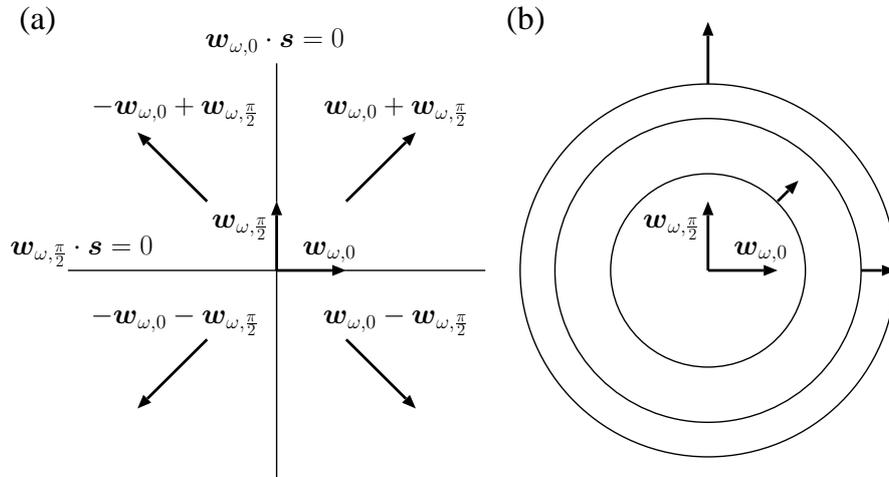


図 2.4: 複雑型細胞モデルの応答特性. (a) $n_i = 1$ の場合. (b) $n_i = 2$ の場合. 同心円は応答の等高線を模式的に示し, 円周から延びた矢印は, その位置での応答の勾配を示す.

論する. 実験に用いたモデルを図 2.3 に示す. その他のパラメータについては, $i \in \{0, 1, 2, 3\}, \phi_i = \frac{\pi i}{2}, a_i = 1, n = 1$ とする.

$n_i = 1$ の場合, $R(\mathbf{x})$ は図 2.4 のように二つの超平面で区切られた区分線形関数である. ここで, $\mathbf{w}_{\omega,0} = -\mathbf{w}_{\omega,\pi}, \mathbf{w}_{\omega,\pi/2} = -\mathbf{w}_{\omega,3\pi/2}$ は局所フーリエ変換の基底に相当する. したがって, これら 4 つのベクトルに張られる二次元平面は, 入力画像のフーリエ係数 $\mathcal{F}[\mathbf{x}](\omega_x, \omega_y)$ を表すガウス平面と見ることができる. $\mathbf{w}_{\omega,\phi_i}$ との内積によって, \mathbf{x} はこの平面に投影されることになるので, このモデルの応答特性についてはこの平面上で議論することができる. 各領域に示したベクトルは, その領域の局所線形カーネルを表す. また, これらの方向は位相を表している. 例えば, 図 2.4 の右上の領域は, 空間周波数 ω の位相が $0 \sim \frac{\pi}{2}$ となる入力画像の集合を表しており, この領域における局所線形カーネル $\mathbf{w}_{\omega,0} + \mathbf{w}_{\omega,\pi/2}$ の位相は $\frac{\pi}{4}$ である. 以上の議論から, 前節で述べた方法を用いて局所線形カーネルを推定すると, 単純型細胞の受容野形状から位相が 45° ずれた, 4 種類の画像が得られると予想できる.

図 2.5 は, 本手法で求めた \mathbf{m} の周りの局所線形カーネル $\hat{\mathbf{x}}$ の例である. \mathbf{m} は分散 0.3 のホワイトガウスノイズである. また, \mathbf{m} に重畳したノイズは分散 0.1 のホワイトガウスノイズである. \mathbf{m} の値によって平均発火率が異なるため, 収束の度合いが異なっているが, 方位と空間周波数が共通で位相が異なる 4 種類のガボールパッチ状のパターンが得られた. \mathbf{m} に含まれる ω 成分の位相と $\hat{\mathbf{x}}$ の位相の関係を図 2.6(a) に示す. 横軸の $0^\circ \sim 90^\circ$ の範囲は, 図 2.4 の右上の領域に相当する. \mathbf{m}

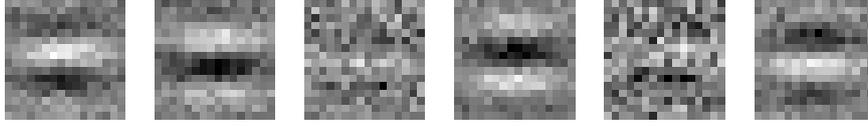


図 2.5: 局所線形カーネルの推定結果.

がこの領域にあるとき、 w_i の位相は 45° 付近で一定である。また他の領域についても同様である。したがって、図2.4のような局所線形カーネルが正しく推定できていることが確認できる。

$n_i = 2$ の場合も同様に、モデルの応答特性を図2.4の平面上で考えることができる。この場合、 $R(\mathbf{x})$ の値は原点からの距離の二乗に比例して大きくなる。したがって、図2.4(b)の矢印のように、 \mathbf{m} の周りの局所線形カーネルは、 \mathbf{m} が持つ空間周波数 ω の成分に等しい位相を持つことが予想される。実際に、 $n_i = 1$ の場合と同様の方法で局所線形カーネル $\hat{\mathbf{x}}$ を求めると、図2.6(b)のように、 \mathbf{m} と $\hat{\mathbf{x}}$ の位相が等しいことが確認できる。

以上より、本手法を用いて、線形逆相関法では得られない複雑型細胞の特性が、複数の線形フィルタ特性として得られることが確認された。また、二つの実験結果から、機能的構成要素の非線形性 n_i の違いが、得られる局所線形カーネルの形状に反映されることが示された。実際の複雑型細胞では、個々の構成要素毎に n_i が異なる場合や、より多くの構成要素を持つ場合が考えられる。

これまでにも様々な研究で、複雑型細胞の応答特性を分析する手法が提案されている。たとえば、スパイクを生じさせた刺激 \mathbf{x} の要素間の二次の関係を調べる手法がある[81, 78]。図1.8のように、スパイク k に先行する刺激を $\mathbf{x}_k(k = 1, \dots, K)$ とすると、この手法では以下の計算を行う。

$$\mathbf{u}_i = \frac{1}{K} \sum_{k=1}^K x_{ki} \mathbf{x}_k \quad (2.12)$$

これを subunit と呼ぶ。ここで、 s_{ki} は \mathbf{x}_k の i 番目の要素を表す。 \mathbf{u}_i は、ニューロンに興奮的に作用する s_i と他の画素との関係を表している。 s_i が表す画素の位置を参照位置と呼ぶ。複雑型細胞の subunit は、参照位置を中心とするガボールパッチ状の形状であることが知られている。Pack ら [81] は、参照位置による subunit の形状の違いはないとして、subunit の平均を求めて議論している。一方 Sasaki ら [78] は、参照位置による違いなどについてより詳しい考察を行っている。

subunit は参照位置を中心とするガボールパッチ状の形状であるため、受容野の中心を基準とした位相は参照位置の座標によって決まる。 $n_i = 1, n_i = 2$ の二つの

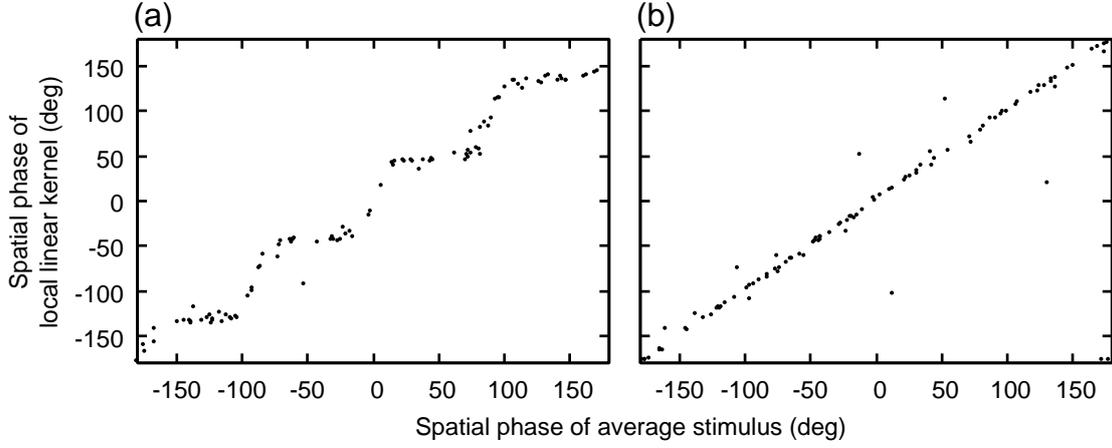


図 2.6: m とその周りの局所線形カーネル \hat{x} の位相の関係. (a) $n_i = 1$ の場合. (b) $n_i = 2$ の場合.

場合について、参照位置と subunit の位相の関係をグラフにしたものを図 2.7 に示す. このグラフからわかるように、subunit の位相は n_i の違いを反映していない. これは、 n_i の値に関わらず、モデルニューロンの偶対称な応答特性が二次の非線形性として検出されるためと考えられる. 多項式近似において、 $|g|$ と g^2 の違いを調べるためには、4 次以上のカーネルを求める必要がある.

2.4.2 方位交差抑制モデル

前節の数値実験で用いたモデルは、興奮性のフィルタ特性のみで構成されている. しかし、実際の視覚皮質では、刺激に含まれる最適刺激とは異なる成分が抑制的に作用するニューロンが存在する. Heeger ら [82] は、このような抑制的な作用を考慮した単純型細胞と複雑型細胞のモデルを提案している. 彼らの単純型細胞モデルの応答特性を次式に示す.

$$\begin{aligned}
 R(\mathbf{x}) &= k \frac{[\mathbf{w}_{i,\phi} \cdot \mathbf{x}]_+^2}{\sigma^2 + \sum_j \sum_{\psi \in \{0,90,180,270\}} [\mathbf{w}_{j,\psi} \cdot \mathbf{x}]_+^2} \\
 &= k \frac{[\mathbf{w}_{i,\phi} \cdot \mathbf{x}]_+^2}{\sigma^2 + \sum_j \sum_{\psi \in \{0,90\}} (\mathbf{w}_{j,\psi} \cdot \mathbf{x})^2}
 \end{aligned} \tag{2.13}$$

ここで、 i, j は方位と空間周波数の組み合わせを表し、 ϕ, ψ は空間位相を表す. また、 k は応答の範囲を決定する正の定数である. このモデルは、単純型細胞に類似する単一の興奮性要素と、エネルギーモデルに類似する複数の抑制性要素で構成

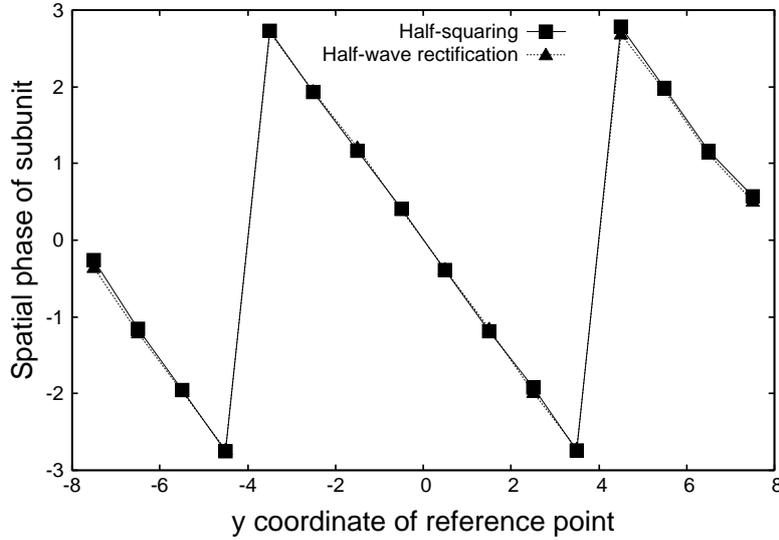


図 2.7: 参照位置の y 座標と得られた subunit の空間位相の関係。

されている。しかし、線形逆相関法を用いると、興奮性要素の持つ線形カーネルのみが得られる。図 2.8 はこのモデルの応答特性を模式的に示している。左右の曲線は、抑制信号により + で示される発火頻度の高い領域が狭くなっていることを表す。抑制性要素がエネルギーモデルのように偶対称のとき、様々な方向から一様に抑制されるため、STA は + の領域の中央に収束する。これは、抑制性要素が存在しない場合の結果と同様である。したがって、線形逆相関法で求めたフィルタ特性には、以上のモデルのような抑制的な側面は現れない。

このモデルの応答特性 $R(\mathbf{x})$ の勾配は以下のように求められる。

$$\nabla R(\mathbf{x}) = 2kA(\mathbf{w}_{i,\phi} - A \sum_j \sum_\psi (\mathbf{w}_{j,\psi} \cdot \mathbf{x}) \mathbf{w}_{j,\psi}), \quad (2.14)$$

$$A = \frac{\mathbf{w}_{i,\phi} \cdot \mathbf{x}}{\sigma^2 + \sum_j \sum_\psi [\mathbf{w}_{j,\psi} \cdot \mathbf{x}]_+^2} \quad (\mathbf{w}_{i,\phi} \cdot \mathbf{x} > 0)$$

ただし、 $\mathbf{w}_{i,\phi} \cdot \mathbf{x} \leq 0$ のとき $\nabla R(\mathbf{x}) = 0$ である。 k, A は非負のスカラーであるため、この式はモデルを構成する線形フィルタの重み付き和を表している。したがって、局所線形カーネルを求めることで、 $R(\mathbf{x})$ を構成する線形フィルタ特性が得られることが期待できる。また、 \mathbf{x} が $\mathbf{w}_{j,\psi}$ に類似している場合 ($\mathbf{w}_{j,\psi} \cdot \mathbf{x} > 0$)、 $\mathbf{w}_{j,\psi}$ の係数が負になることがわかる。よって、除算的な抑制を線形カーネルの減算的な重ね合わせとして取り出すことができると考えられる。

以上を確認するため、一つの抑制性要素を持つ単純なモデルを用いて数値実験を行った。図 2.9 に示すように、このモデルは縦の方位に選択的な興奮性要素 $\mathbf{w}_{v,0}$

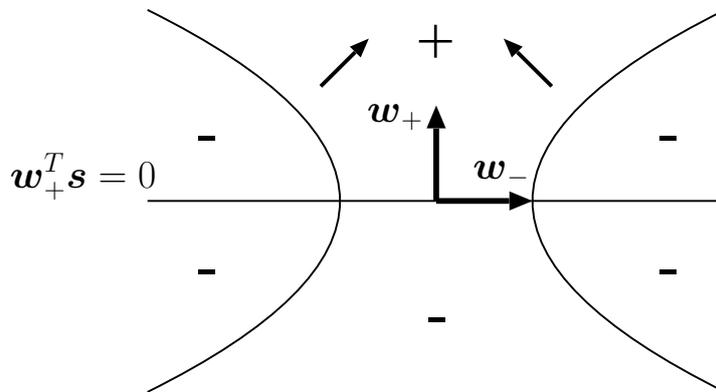


図 2.8: 抑制モデルの応答特性. w_+, w_- はそれぞれ興奮性と抑制性の線形フィルタを表す. 発火頻度の高い領域を+, 低い領域を-で示す. また斜めの矢印はその位置の勾配を表す.

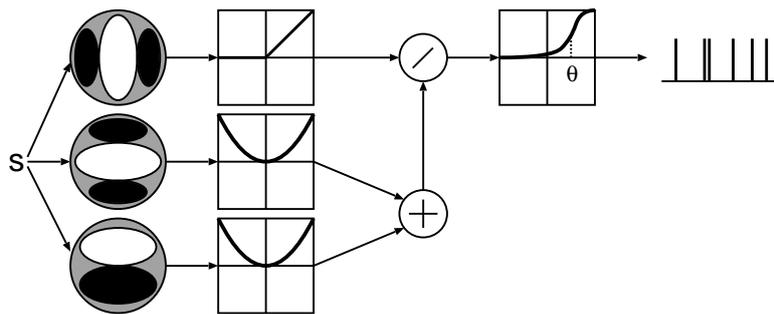


図 2.9: 数値実験に用いた抑制モデル.

と, 横の方位に選択的な抑制性要素 $w_{h,0}, w_{h,\frac{\pi}{2}}$ からなる. 抑制性要素は位相に選択性を持たない. 得られた局所線形カーネルの例を図 2.10(a) に示す. これらは縦と横の縞模様を重ね合わせたパターンと見ることができる. よって上で述べた通り, $w_{v,0}, w_{h,0}, w_{h,\frac{\pi}{2}}$ を重ね合わせたフィルタ特性が局所線形カーネルとして得られた. これらが直交する場合, 次式で局所線形カーネル \hat{x} から $w_{v,0}$ の成分を取り除くことができる.

$$\hat{x}' = \hat{x} - \frac{w_{v,0} \cdot \hat{x}}{|w_{v,0}|^2} w_{v,0} \tag{2.15}$$

$w_{v,0}$ には線形逆相関法の結果を用いることができる. この処理を行った結果を図 2.10(b) に示す. この結果より, 抑制性要素が位相に選択性を持たないことを確認できる.

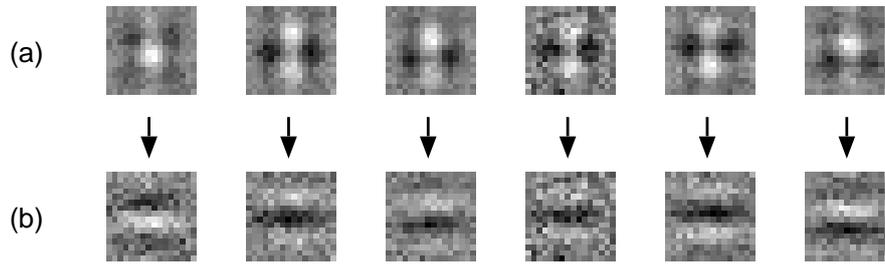


図 2.10: 抑制モデルの局所線形カーネルの推定結果. (a) 局所線形カーネル. (b)(a) から従来の逆相関法の結果として得られる線形カーネルの成分を取り除いたもの.

以上より、本手法を用いて除算的な抑制性要素を持つニューロンの特性を分析することができる。ここで用いたモデルが持つ抑制性要素は一つであるため、式 (2.15) を用いて興奮性要素の特性を取り除くことで、横線分が抑制的に作用することを確認することができた。しかし、Heeger らのモデルのように、より多くの抑制性要素を持つニューロンに本手法を用いた場合、得られる局所線形カーネルには、様々な空間周波数選択性を持つ線形フィルタ特性が混在することになる。これらが直交する場合、主成分分析を用いることが考えられるが、少なくともここで線形フィルタとして用いたガボール関数は互いに直交しない。このため、得られた複数の局所線形カーネルから単一の線形フィルタの形状を引き出すためには、さらなる検討が必要である。

2.5 まとめ

本章では、逆相関法に区分線形近似を導入し、非線型なニューロン応答の局所的な特性を分析する手法を提案した。この手法の利点として、まず結果の表現が簡便であることが挙げられる。後述するように、ニューロン応答特性を多項式で近似する方法が考えられる [83]。しかしニューロンの非線型性が強い場合、完全な同定のためには高次の項までの分析が必要である。これは収束性の観点から困難である。また、入力が多変量であるため、組み合わせ爆発により、推定すべきカーネルが非常に多くなるという問題がある。本手法ではこの問題に対し、入力空間における局所領域を対象に測定し線型近似することによる解決を試みている。

局所線型カーネルは、ニューロンの入出力特性 $F(\mathbf{x})$ の微分に近似的に等しい。これを適応的な分析に応用することが考えられる。例えば、 $F(\mathbf{x})$ の勾配を推定しながら測定する領域を移動させて行き、ニューロンの最適刺激を探索することができる。また、最適刺激付近を測定することで、視覚刺激に含まれる複数の特徴

がどのように相互作用するのかを調べられることが期待できる。

ニューロン応答特性は，呈示刺激の空間 \mathcal{X} 上の関数として表すことができる．この \mathcal{X} 全体における応答特性を推定することが，常に必要であるとは限らない．例えば，二つの線分の位置関係に対してニューロン応答がどのような特性を持っているのかを調べるのが目的であれば，各線分の位置や方位を表す数個のパラメータ θ で表現される画像を呈示し，入力と θ の関係を分析すればよい．このような状況を記述するため， \mathcal{X} を

$$\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \quad (2.16)$$

のように，関心のあるパラメータで表現される部分空間 \mathcal{X}_1 とそれに直交する部分空間 \mathcal{X}_2 の直積で表し， $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{X}_1 \times \mathcal{X}_2$ とおく．subspace-RC[84] では，選択された基底が張る空間が \mathcal{X}_1 に相当し， $\mathbf{x}_2 \in \mathcal{X}_2$ は $\mathbf{0}$ に固定されるので，呈示刺激の分布は $p(\mathbf{x}) = p(\mathbf{x}_1)\delta(\mathbf{x}_2)$ となる．ニューロン応答特性を $F(\mathbf{x}) = F(\mathbf{x}_1, \mathbf{x}_2)$ とおくと，subspace-RC で推定される特性は $F(\mathbf{x}_1, \mathbf{0})$ となる．

一方，本手法では測定する領域を \mathcal{X} の部分空間ではなく，より一般の部分集合として選択することになる：

$$\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2. \quad (2.17)$$

また，推定される特性は $F(\mathbf{x})|_{\mathcal{X}_1}$ となる．従って，subspace-RC よりも柔軟であるが，どの領域を測定するかを決定するためにより多くの仮定を要する．

第3章 階層ベイズモデルによる組み合わせ表現の獲得

3.1 目的

V1の方位選択性を持つニューロンは、空間位相に選択性を持つ単純型細胞と、これを持たない複雑型細胞の二種類に分けることができる [18, 19]. これまでの様々な生理学的実験によって、これらのニューロンの詳細な応答特性に関する知見が得られており (例えば [2]), 機能的に等価なモデル (複雑型細胞のエネルギーモデルなど) も提案されている.

しかし、この機能がどのように獲得されているのかという疑問を解決することも必要である. 1.3.4節で述べた通り、V1における受容野の自己組織的獲得を生成モデルの観点から説明する、スパースコーディング (SC) 仮説が提案されている. SCは単純なベイズモデルで記述できる. いま、各時刻の入力パターン \mathbf{x}_t を行列 $X = (\mathbf{x}_1, \mathbf{x}_2, \dots)$ にまとめて表し、基底のセットを行列 A 、応答 \mathbf{y}_t を入力パターンと同様に行列 U で表す. このとき上記の問題は

$$X = AU + N \quad (3.1)$$

と定式化される. ここで N は平均0のホワイトガウスノイズを表す. また A, U の事後分布は

$$p(A, U|X) \propto p(X|A, U)p(U) \quad (3.2)$$

となる. ここで U の事前分布 $p(U)$ としてコーシー分布やラプラス分布などのスパースな分布を導入したもの (Aについては何らかの行列ノルムに制約を課すことが多い) がSCであると考えられる.

尤度 $p(X|A, U)$ が A, U をパラメータとする確率分布であるのと同様に、 U の事前分布 $p(U)$ もまたパラメータを持つ確率分布である. 例えばコーシー分布やラプラス分布は尺度母数を持っている. この事前分布のパラメータを確率変数と考え、その生成モデルを導入したものが階層ベイズモデルである.

Karklin ら [59, 60] のモデルでは、この階層ベイズに基づいて、自然画像からより高次の統計的構造を抽出することに成功している. 彼らのモデルでは、 $p(U)$ の

尺度母数が非負の非線形関数 ϕ と行列 B, V で $\phi(BV)$ と表される. このとき, 式 (3.2) の右辺は $p(X|A, U)p(U|B, V)p(V)$ となる. 同様に階層ベイズに基づくモデルに Topographic ICA (TICA) がある [56, 57]. このモデルでは, B はニューロンの近傍関数を定義する固定値として定義される. これにより, 近傍のニューロンは応答の絶対値に正の相関を持ち, 学習後に類似する基底を獲得することになる. TICA はこのような学習機構によって V1 のハイパーコラム内の方位選択性マップを再現することに成功している. 加えて, このモデルの V は複雑型細胞の応答を示すことになるため, TICA は複雑型細胞の応答特性の獲得について統計的な説明を与えるものと考えることができる.

しかしながら, これらの階層ベイズモデルには, 複雑型細胞が単純型細胞から偶対称な入力を受けるという, 強い仮定に基づいているという問題がある. これは, 単純型細胞の応答を負値を取りうる行列 U で表し, その絶対値を上位層で表現しているためである. 時刻 t におけるニューロン j の応答 u_{jt} が負値を取るとは, 皮質上では \mathbf{a}_j (A の j 番目の列ベクトルを表す) を反転させた受容野を持つニューロンが強度 $-u_{jt}$ で応答していることに相当する. 従って, この u_{jt} の絶対値を取ることにより, 逆位相の受容野を持つ二つのニューロンからの等しい結合を暗に仮定していることになる.

以上より, 複雑型細胞-単純型細胞間の偶対称な結合がどのように獲得されているのかが問題となる. 本章では, 階層ベイズモデルの第一層に非負性を導入することにより, 単純型細胞応答を模したモデルを構築し, このような結合を自然画像の統計的性質から獲得できる可能性について検討する [85, 86].

3.2 モデル

本節では, ニューロンの非対称な応答特性を反映した階層ベイズモデルを構築し, 視覚野ニューロンとの対応について議論する. ここでは, 網膜から V1 に至る経路を, 双方向結合を持つ二階層のネットワークとしてモデル化する. 外界からの入力を列ベクトル \mathbf{x} , V1 のニューロン集団の応答を二つの列ベクトル \mathbf{u} , \mathbf{v} で表す. \mathbf{u} は単純型細胞, \mathbf{v} は複雑型細胞に対応する. 以下, 特定の入力パターンやそれに対する応答を指す場合は \mathbf{x}_t , \mathbf{u}_t のように添字をつけて表現する. また, 入力や応答の全体は, 列ベクトルを横に並べた行列として U のように大文字で表す. 網膜と V1 の単純型細胞との入出力関係は, 線形フィルタとその応答の非線形関数で近似することができる [2]. そこで, この線形受容野を行列 A で表す. A の各列は単純型細胞の受容野に対応する.

まず, 第一層は下式のように, A の基底ベクトルの線形和により入力パターン

の線形予測を行っているとは仮定する.

$$\mathbf{x} = A\mathbf{u} + \mathbf{n} \quad (3.3)$$

ここで $\mathbf{n} \sim N(0, \sigma_x^2)$ はホワイトガウスノイズを表す. \mathbf{u} は発火頻度を表している
ので, 非負の確率変数とする.

また SC が示す通り, 空間的に局在する基底を獲得するためには, \mathbf{u} の活動はス
パースである必要がある. そこで, \mathbf{u} の事前分布を指数分布とする. 指数分布の
確率密度関数は以下の式で表される.

$$p(x) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right) \quad (3.4)$$

θ は指数分布の尺度母数であり, 期待値と分散に等しい. この尺度母数が \mathbf{v} によっ
て生成されるという以下のモデルを導入する.

$$p(\mathbf{u}|B, \mathbf{v}) = \prod_i \frac{1}{\theta_i} \exp\left(-\frac{u_i}{\theta_i}\right), \quad (3.5)$$

$$\boldsymbol{\theta} = \theta_0 \exp(B\mathbf{v})$$

式 (3.5) は第一層からみると \mathbf{u} の事前分布である. \mathbf{v} の活動により θ_i が大きくな
れば, u_i の値が平均的に増大する. また θ_i が小さくなれば, u_i の値が小さくなり,
相対的にスパースに分布することになる. ただし, θ_i は指数関数によって抑えら
れているため, 0 になることはない.

逆に第二層からみると, 式 (3.5) は \mathbf{v} の尤度関数となっている. 式 (3.5) の対数
をとって v_j で偏微分すると

$$\frac{\partial}{\partial v_j} \log p(\mathbf{u}|B, \mathbf{v}) = - \sum_i B_{ij} \left(\frac{u_i}{\theta_i} - 1 \right) \quad (3.6)$$

が得られる. この式より, u_i の値とその予測である θ_i の比が 1 になるように \mathbf{v} が
変化することが分かる.

最後に, \mathbf{v} の事前分布として次式のコーシー分布を用いた.

$$p(v_j) = \left(\pi\gamma \left(1 + \frac{v_j^2}{\gamma^2} \right) \right)^{-1} \quad (3.7)$$

γ はコーシー分布の尺度母数である. $\mathbf{v} = \mathbf{0}$ のとき, $\boldsymbol{\theta}$ の各要素は θ_0 となる. 従っ
て, この事前分布は $\boldsymbol{\theta}$ の各要素が θ_0 を中心に変化するような制約条件を課してい
ることになる. 以上をまとめると, 本モデルは図 3.1 のような生成モデルとなる.

以下では, 偶対象な非線形性を仮定せずに複雑型細胞の応答特性を獲得するこ
とが困難であることを示し, これを本モデルがどのように解決し得るかについて

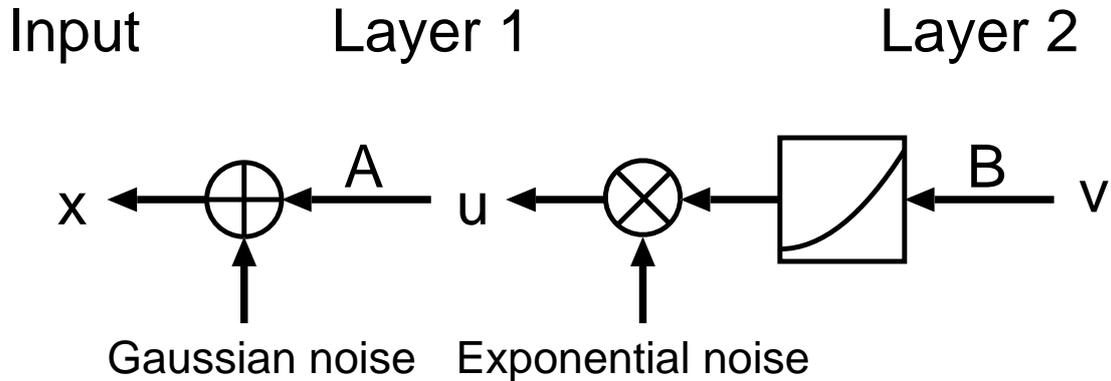


図 3.1: 本研究で考える二階層生成モデル.

議論する. 同じ空間周波数選択性を持つ単純型細胞のうち, 空間位相が近いものの応答は正の相関を持つと考えられる. 一方, 互いに逆位相のものでは負の相関を持つ. このような相反する統計的性質を持つ特徴を, 単一のモデルでどのように統一的に表現するのが問題となる.

期待値 θ の指数乱数は, 期待値 1 の指数乱数に θ をかけたものに等しい. 従って, 第二層のモデルは $\mathbf{y} = \boldsymbol{\theta} \otimes \mathbf{n}_e$ (ここで \mathbf{n}_e は期待値 1 の指数乱数, また \otimes はベクトルの要素毎の積を表す) となるように \mathbf{y} の近似を行っているともみることができる. これは予測誤差として指数乱数が乗法的に重畳されることを意味している. このため, 指数分布の形状を考慮すると, このモデルは u_i の実際の値が予測値 θ_i よりも小さい場合は, その逆の場合と比較して許容しやすいことがわかる.

このことは, 式(3.6)の第 i 項と θ_i , u_i の関係を見ることでも分かる. すなわち, u_i に対して θ_i が小さいときにこの項は u_i の変化に応じて大きく変動するのに対して, u_i が 0 の時には θ_i の値に関わらず, この項の値は高々 B_{ij} である.

以上より, \mathbf{v} が \mathbf{u} に対して論理和のように応答するということができる. つまり, v_j との結合を持つ複数の u_k ($k \in \{i \mid B_{ij} \gg 0\}$) のうち, 一部が 0 に近い値を取っていても, v_j の値が大きく減少することはないと考えられる. これは, \mathbf{v} が複雑型細胞に対応する空間位相不変性を獲得する上で重要な要素であるといえる.

さらに, 第一層と第二層の相互作用について, 以下のように考えることで, 位相不変性の獲得の可能性を示唆することができる. 元々類似する基底を持つ $u_i, u_{i'}$ は, 共通の v_j に対して大きい結合を持つ可能性が高い. このとき, u_i の基底にほぼ相似なパターンが入力され, u_i の活動が大きくなると, v_j を介して $u_{i'}$ の活動も大きくなる. 従って, 一時的に u_i の周辺のみスパースネスが緩和された状態になると考えられる. ここで u_i とは大きく異なるが $u_{i'}$ には近い基底を持つ $u_{i''}$ の活動

が予測よりも大きくなることで、結合 $B_{i''j}$ が強化され、 $u_{i''}$ と u_i との関係性が $u_{i'}$ を介して v_j の基底に符号化されることになる。このように、第一層と第二層が互いの活動を強化し合う機構によって、第一層の活動には反映されない複数の特徴の関係性が第二層によって記述されることが期待できる。

第一層のモデルを行列形式で表すと、 $X \approx AU$ のように非負制約付きの行列の分解とみなせる。これは行列分解の手法である NMF (non-negative matrix factorization) [87, 88] に類似している。NMF は、入力と混合行列を含めて全ての行列に非負性を仮定した手法であり、 X と A に非負性を仮定していない点では異なっている。以下では、この NMF と本モデルとの関係について議論しておく。

負値を含むデータに NMF を適用する際、入力の各次元に対して高さが二倍の非負行列 X' を入力として用いることが多い [89]。この場合、基本ベクトル $e_1, e_2 \dots e_M$ とそれらを反転したベクトルで構成される行列 R で元の入力 $X = RX'$ を表現することができる。この X' に NMF を適用することは、 $X' = SU$ のように定式化される。ここで S は NMF の出力となる非負の混合行列を表す。第一層のモデルの学習は、この正負の分離と非負行列による線形変換を合成した $A = RS$ を学習することに相当する。また学習で得られた A に対して X と同様の操作を行うことで S が得られる。

以上より、入力に負値を含む本モデルの場合は、 U のみに非負制約を課すことで NMF と同等の機能を実現できると考えられる。ただし、ここでの議論は第一層のみを考えた場合に成り立つ。上で述べた通り、実際は U は上位層からの信号によって修飾されるため、結果として得られる応答や基底は異なると考えられる。

次節では上記のモデルに基づいて自然画像の学習を行い、獲得された特性について議論する。

3.3 勾配法による MAP 推定

ここでは、勾配法を用いてモデルパラメータの最大事後確率 (MAP) 推定を行う。モデルパラメータの最適化は、以下の目的関数の最小化問題に帰着する。

$$E = -\log p(\mathbf{x}|A, \mathbf{u})p(\mathbf{u}|B, \mathbf{v})p(\mathbf{v}) \quad (3.8)$$

これを実行するため、各パラメータを乱数で初期化し、目的関数 E の各パラメータに関する勾配を求め、適当なステップ幅で各パラメータの更新を行った。入力画像には自然画像から抽出した 16×16 ピクセルのパッチを用いた。各パッチは平均と分散がそれぞれ 0, 1 になるように正規化した。前述の通り、 \mathbf{u} には非負制約が課されている。これを実現するため、 $\mathbf{u} = \exp(\mathbf{u}')$ となる \mathbf{u}' を導入し、 \mathbf{u} の代

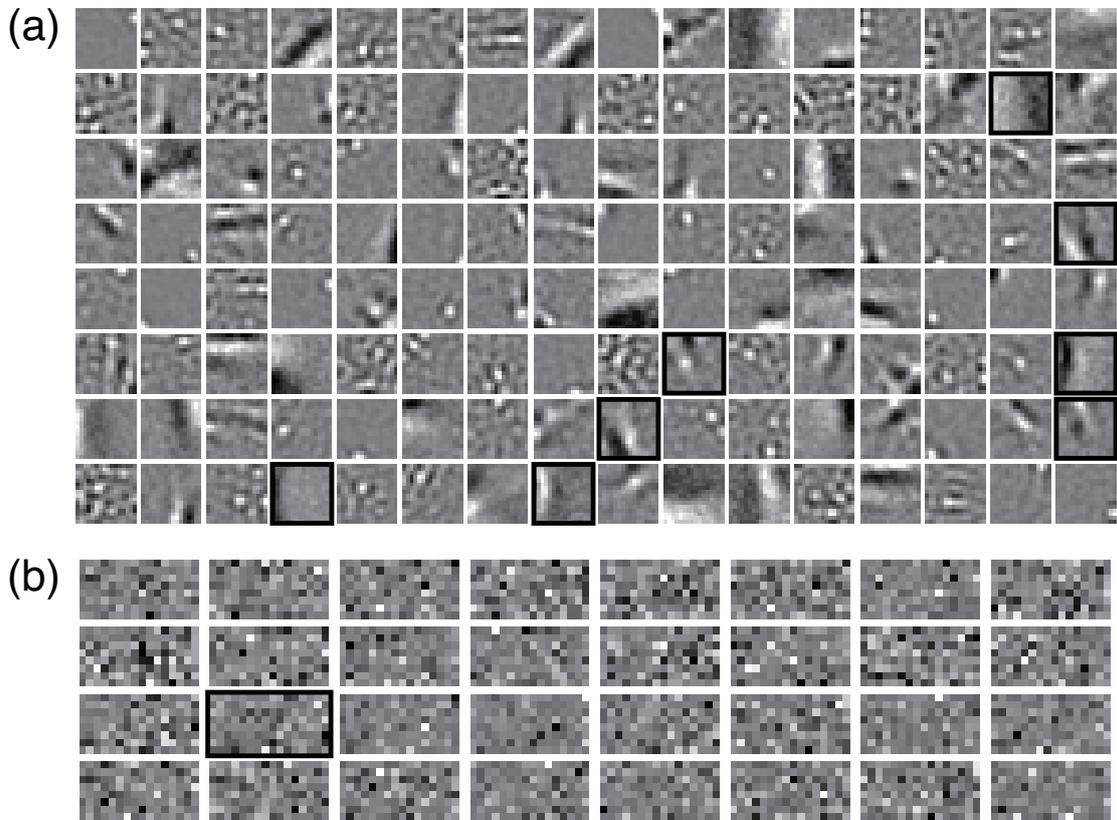


図 3.2: 自然画像の学習によって獲得された基底. 明るいピクセルと暗いピクセルはそれぞれ正と負の結合を表す. (a) 第一層. (b) 第二層. 各ピクセルの配置は (a) での基底の配置と対応している.

わりに \mathbf{u}' を更新した. \mathbf{u} と \mathbf{v} の次元はそれぞれ 128, 32 とした. また, モデルのハイパーパラメータである σ_x^2 , θ_0 と γ はそれぞれ 0.005, 0.01, 0.1 に設定した.

学習で獲得された基底を図 3.2 に示す. (a) と (b) はそれぞれ A と B の基底を並べたものであり, 各基底の明るい点と暗い点はそれぞれ下位層との正と負の結合を表す. (a) を見ると, 第一層では従来の SC と同様の基底が得られていることが分かる. ただし, \mathbf{u} に非負制約があるため, 従来の SC のように一つの基底でその逆位相のパターンをも表現することはできない. この制約条件下で予測誤差を最小にするには, 各受容野位置, 空間周波数と方位に対して, 逆位相, もしくはいくつかの位相に対応する基底を個別に獲得しなければならない. 実際にこのような組み合わせがいくつか存在している. 例えば, 図 3.3 右の \mathbf{a}_{64} と \mathbf{a}_{105} は互いに逆位相のガボールパッチを表現している.

図 3.3 は第二層のニューロン v_{18} と \mathbf{u} との結合をより詳細に示したものである.

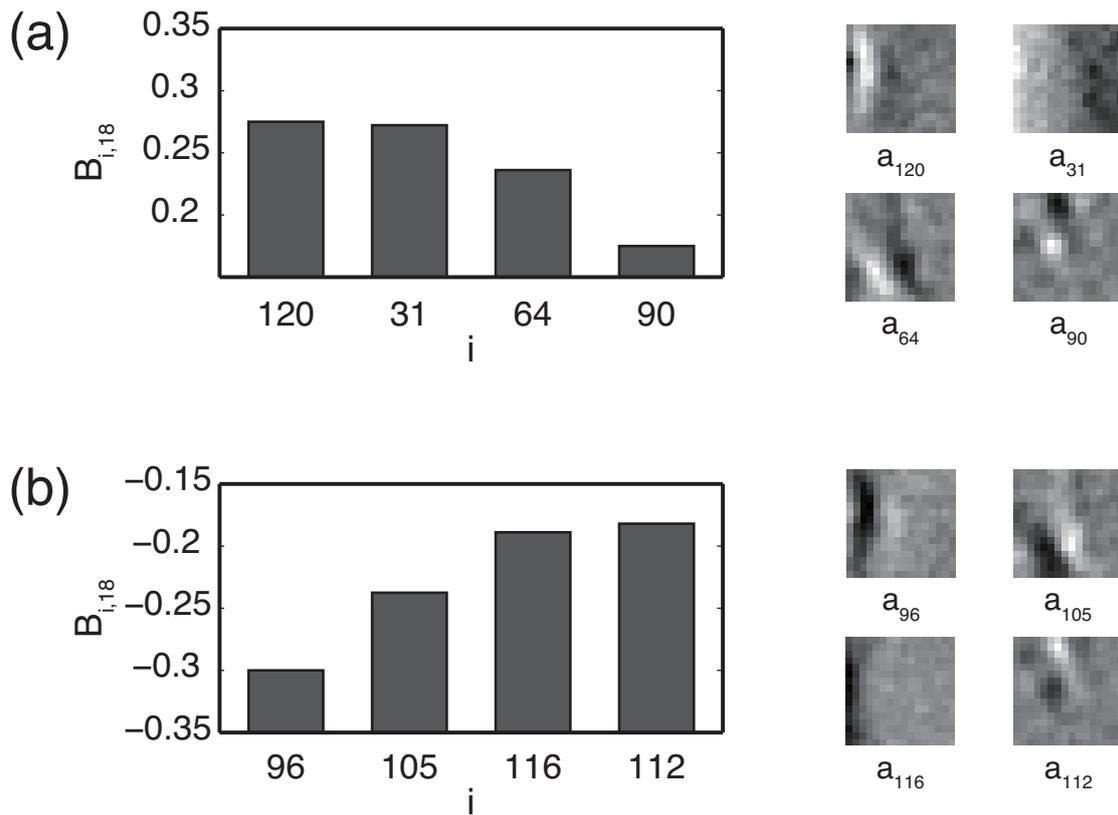


図 3.3: 第二層のモデルニューロン v_{18} と第一層のモデルニューロンの結合加重. 対応する基底を図 3.2(b) に枠で示す. (a), (b) ではそれぞれ結合加重が大きい順, 小さい順にその値を棒グラフで示し, 右に対応する第一層の基底を示す (これらは図 3.2(a) において枠で示されている).

(a) と (b) ではそれぞれ正と負の結合が強い順に並べられており, $v_{18} > 0$ のとき, これらの結合は対応する u_i にそれぞれ興奮的, 抑制的に作用する. また, $v_{18} < 0$ のときはその逆が成り立つ. v_{18} に対して同符号の結合を持つ第一層ニューロンの基底を見ると, v_{18} が縦方向に偏った特徴選択性を持っていることがわかる. 特に方位選択性に関して, この様な特性が第二層全体で見られた. この結果より第二層は, 第一層では表現できない特徴のカテゴリを表現することができると考えられる.

前節で述べた通り, v が複雑型細胞の応答特性を獲得するためには, 逆位相を含め, 様々な空間位相特性を持つニューロンとの同符号の結合を獲得する必要がある. しかし, 今回はこのような結合は見られなかった. この理由として, A の基底が少なかったことが挙げられる. 上で述べた位置, 空間周波数と方位が共通するガボールパッチの組み合わせは, 互いに逆位相のものか, 単独で存在するものがほとんどである. これは, 少ない基底で全ての画像を表現しようとしたとき, 空間位相のバリエーションを増やすよりも, 空間周波数や方位のバリエーションを増やした方が効率的であったことを意味している. 従って, 同じモデルのもとで空間位相のバリエーションを増やすためには, 実際の皮質と同様に, 入力次元よりはるかに多い基底数を設定する必要があると考えられる.

また, u がスパース過ぎたことが, 基底の獲得に影響していることも考えられる. 本モデルでは U の行方向と列方向のスパースネスを区別していない. すなわち, 各パターンに反応するニューロンの個数 (population sparseness) と, 各ニューロンが全パターンを通して反応する回数 (lifetime sparseness) を個別に制御するような制約は設けていない. 式 (3.8) から導出される A の更新式には Hebb 則が成り立っているが, これはパターンに対して全く反応しなかったニューロンはそのパターンを学習しないことを意味する. このため, スパースネスが強すぎることで, 全体を通してほとんど学習していないニューロンが存在し, 一部の基底に無視できないノイズが重畳されたものと考えられる.

以上の通り, そもそも第二層が獲得すべき特徴の組み合わせが存在しないことによつて, 期待される結果が得られなかったと考えられる. ただし, もし第一層がそのような基底を獲得した場合に, 第二層が複雑型細胞の特性を獲得できる可能性を示唆する結果も得られた. ガボール関数の特性を考えると, 空間周波数や方位よりも, 空間位相の方が複数の基底の和で表現しやすい. 従って, 第一層が適切な基底を獲得していれば, 前節で述べたような機構により, 様々な空間位相が集約されることで複雑型細胞の特性が獲得されると期待できる.

3.4 改良モデル

前節で示した通り，第二層が複雑型細胞の特性を獲得するためには，第一層の基底を大幅に増やす必要がある．またスパースネスが強すぎるという問題も存在し，これは単に θ_0 の値を変更するだけでは解決できないと思われる．そこで，この節ではこれらの問題を解決するため，新たなモデルを検討する．また，第二層の変数に非負の制約を課し，非線形性についても見直すことで，生理学的により妥当なモデルを構築する．

第一層は前節のモデルと同様，加法性のホワイトガウスノイズを誤差とする線形予測で，尤度関数は以下の通りである．

$$p(\mathbf{x}|A, \mathbf{u}) = \frac{1}{(2\pi\sigma_x^2)^{\frac{M}{2}}} \exp\left(-\frac{1}{2\sigma_x^2}\|\mathbf{x} - A\mathbf{u}\|^2\right) \quad (3.9)$$

また，第二層のモデルとして，指数分布を $x = u^n$ で変数変換した次のような分布を考える．

$$p(\mathbf{u}|B, \mathbf{v}) = \prod_i \frac{n}{\theta_i} u_i^{(n-1)} \exp\left(-\frac{u_i^n}{\theta_i}\right) \quad (3.10)$$

$$\mathbf{u} \geq \mathbf{0}, \quad \boldsymbol{\theta} = \theta_0(\epsilon + B\mathbf{v})$$

ここで $\epsilon > 0$ は定数とし，ベクトルへの加算は要素毎の演算とする． $n > 1$ のとき，この分布関数の減衰は指数分布よりも速くなるので，分布の裾は軽くなる．これによってスパースネスを緩和することができる．

後述するが， n は単純型細胞の static nonlinearity に相当する．このことから，前節のモデルよりも自然な非線形性を第一層-第二層間に導入することができる．ネコのV1では，この n の値が平均的には2.17付近（この実験では標準偏差は0.53だった）であることが分かっている [2]．従って，計算上の都合から最も近い整数を選ぶとすれば， $n = 2$ が望ましいといえる．

式 (3.10) は， n 乗すれば期待値 θ_i の指数分布に従う確率変数の分布である．このため，第二層は次のような近似を行っているといえる．

$$U^n \approx BV \quad (3.11)$$

ここで， n は各成分を n 乗する演算を表している．従って，式 (3.10) を B, \mathbf{v} の尤度関数とみれば，非負行列 U の各成分を n 乗したものに対するNMFと同等である．ただし推定誤差として，指数乱数が乗法的に重畳されることを仮定している．従って，やはり前節で述べたような特性を持つと考えられる．また，式 (3.10) は \mathbf{u} の事前分布でもあるため，実際は二層間の相互作用によって各変数の値が定まる．

このモデルでは \mathbf{v} も \mathbf{u} と同様に非負の確率変数とする。

$$p(\mathbf{v}) = \frac{1}{\theta_v} \exp\left(-\frac{\mathbf{v}}{\theta_v}\right) \quad (\mathbf{v} \geq \mathbf{0}) \quad (3.12)$$

さらに、結合荷重 B も同様に指数分布する非負の変数とする。

3.5 変分ベイズの適用

3.3節の数値実験の結果では、獲得された基底にかなりのノイズが重畳されており、解が不安定であることがわかる。これは学習の各ステップでモデルパラメータの点推定を行っているためであると考えられる。従って、学習結果は初期値や学習中の各ステップでの値に大きく依存することになり、収束が遅いうえに局所解に陥りやすい。

特に第一層の応答は入力層と第二層の両方に依存している。このため、式(3.8)の勾配に基づく更新式では初期値や前ステップでの各変数の値に大きく依存し、収束性や局所解に関して問題がある。従って、各変数の更新には周辺尤度関数を用いるのが理想的である。しかし、モデルを構成する多くの変数に関する周辺化や、EM アルゴリズムを適用するために必要である対数尤度関数の期待値を計算することは困難である。この問題は、視覚皮質が多様な入力に内在する統計的構造をどのように安定的に推定しているのかという、実装の問題にも関わっている。

そこで、学習後のネットワーク特性のより詳細な分析を可能とするため、ここでは上で構築したモデルに変分ベイズ法 [90, 91] を適用することを検討する。変分ベイズ法を用いることで、モデルパラメータのより安定的な求解が望める。また、同じくベイズ推定における計算手法であるマルコフ連鎖モンテカルロ (MCMC) 法とは異なり、変分ベイズでは各変数の事後分布が解析的な表現で得られる。これを利用して、実際の皮質での実装も考慮に入れながらパラメータ推定のアルゴリズムを構築し、これが生理学的に妥当であるかどうかを検証する。

変分ベイズでは、モデルパラメータに関する推定事後分布 $q(\mathbf{u}, \mathbf{v}, A, B)$ を導入し、汎関数

$$\mathcal{F}[q] = \iiint q(\mathbf{u}, \mathbf{v}, A, B) \log \frac{p(\mathbf{x}, \mathbf{u}, \mathbf{v}, A, B)}{q(\mathbf{u}, \mathbf{v}, A, B)} d\mathbf{u}d\mathbf{v}dAdB \quad (3.13)$$

を q について最大化することにより、 q と真の事後分布との KL ダイバージェンスを間接的に最小化する。ここでは

$$q(\mathbf{u}, \mathbf{v}, A, B) = \prod_j q(\mathbf{u}_j) \prod_k q(\mathbf{v}_k) \prod_{i,j} q(A_{ij}) \prod_{j,k} q(B_{jk}) \quad (3.14)$$

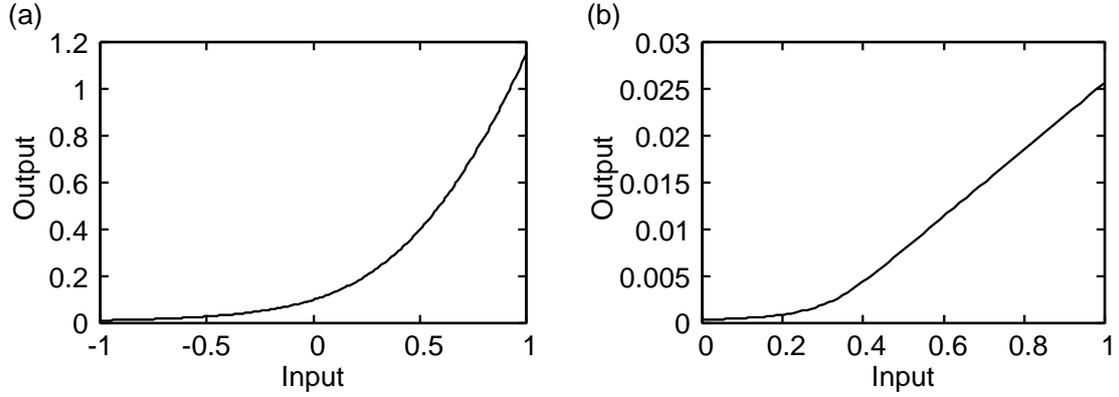


図 3.4: 変分ベイズ法により導出された (a) u , (b) v の入出力関数.

のように、全ての変数が独立である事後分布によって真の事後分布の近似を試みる.

あるモデルパラメータ z の推定事後分布 $q(z)$ に関する $\mathcal{F}[q]$ の停留点を求めるために $\frac{\delta \mathcal{F}[q]}{\delta q(z)} = 0$ とおくと,

$$q(z) \propto \exp \langle \log p(\mathbf{x}, \mathbf{u}, \mathbf{v}, A, B) \rangle_{q(\{\mathbf{u}, \mathbf{v}, A, B\} \setminus z)} \quad (3.15)$$

が得られる. ここで $\langle \cdot \rangle_{q(\{\mathbf{u}, \mathbf{v}, A, B\} \setminus z)}$ は z 以外の変数に関する期待値を表す. $q(z)$ を求めるために他の変数の事後分布が必要であるため, この変分問題を閉形式で解くことはできない. そこで各変数の事後分布のパラメータを適当な初期値で初期化し, 全ての変数について式 (3.15) による更新を交互に繰り返すことで最適な事後分布を推定する. 以下ではモデルの各変数の推定事後分布を導出し, ネットワーク特性について解析的に議論する.

まず, u_j の事後分布の更新式は以下ようになる.

$$q(u_i)^{(t+1)} \propto \exp \left(-\frac{(u_i - \mu_{1i}^{(t)})^2}{2(\sigma_{1i}^{(t)})^2} \right) u_i^{n-1} \exp \left(-\frac{u_i^n}{\langle \theta_i \rangle^{(t)}} \right) \quad (u_i \geq 0) \quad (3.16)$$

$$\mu_{1i}^{(t)} = \left\langle \frac{\mathbf{a}_i \cdot \mathbf{x} - \sum_{i' \neq i} \mathbf{a}_i \cdot \mathbf{a}_{i'} u_{i'}}{\|\mathbf{a}_i\|^2} \right\rangle^{(t)} \quad (3.17)$$

$$(\sigma_{1i}^{(t)})^2 = \left\langle \frac{\sigma_x^2}{\|\mathbf{a}_i\|^2} \right\rangle^{(t)} \quad (3.18)$$

t は学習のステップを表す. 最初の指数関数が尤度, 残りの部分が事前分布に対応する. 式 (3.16) では $\langle x^{-1} \rangle \approx \langle x \rangle^{-1}$ のような近似を用いている. $\mu_{1j}^{(t)}$ は受容野 \mathbf{a}_j からの入力と他のニューロンからの抑制性で構成されている. また, 抑制性

の入力は基底の内積 $\mathbf{a}_j \cdot \mathbf{a}_{j'}$ で重み付けされている。従って、全く異なる基底を持つニューロン間には相互作用がないことがわかる。V1で見られる網膜部位再現性 (retinotopy) を考慮すると、これは受容野位置が離れた細胞間の水平結合がほとんど存在しない [92] ことに対応する。また、 $n = 2$ の場合、指数関数をまとめると

$$q(u_i)^{(t+1)} \propto u_i \exp\left(-\frac{(u_i - \mu_{2i}^{(t)})^2}{2(\sigma_{2i}^{(t)})^2}\right) \quad (3.19)$$

$$\mu_{2i}^{(t)} = \frac{\langle \theta_i \rangle^{(t)}}{\langle \theta_i \rangle^{(t)} + 2(\sigma_{1i}^{(t)})^2} \mu_{1i}^{(t)} \quad (3.20)$$

$$(\sigma_{2i}^{(t)})^2 = \frac{\langle \theta_i \rangle^{(t)}}{\langle \theta_i \rangle^{(t)} + 2(\sigma_{1i}^{(t)})^2} (\sigma_{1i}^{(t)})^2 \quad (3.21)$$

と書ける。式 (3.20) と式 (3.21) の分数の部分は、 $\langle \theta_i \rangle$ に対して 0 から 1 の値を取る単調増加の関数になっている。よって、第二層からの予測信号が \mathbf{u} のゲインをコントロールするという構造になっていることがわかる。

\mathbf{v} の事後分布も同様にして求めることができる。ただし、式 (3.16) と同様の近似を用いる。 \mathbf{v} の事後分布を求めるためには、 $\langle u_i^2 \rangle^{(t)}$ が必要になる。よって、 u_i の事後分布 $q(u_i)$ の二次モーメント $\langle u_i^2 \rangle$ が \mathbf{v} に出力されると考える。この二次モーメントは、不完全ガンマ関数を用いて $\mu_{2j}^{(t)}, (\sigma_{2j}^{(t)})^2$ の関数として記述できる。この関数を $\mu_{2j}^{(t)}$ についてプロットしたものが図 3.4(a) である。これを見ると、 u_j が入力信号 $\mu_{2j}^{(t)}$ に対して、Anzai ら [2] が測定したネコの V1 における単純型細胞の非線形性に類似する入出力特性を持つことがわかる。また、 $\langle \mathbf{u}^2 \rangle$ の入力に対する v_j の入出力関数は、第一層-第二層間の結合荷重 B の分布によって異なるが、図 3.4(b) のような形状になる。

以上をまとめると、本モデルに変分ベイズを適用することで、図 3.5 のようなアルゴリズムが導出される。第一層は各ニューロン毎に受容野からの入力と水平結合による抑制信号を受け、これが第二層からの乗法的な興奮性信号によってゲインコントロールされた後、非線形な入出力関数を経て第二層に出力する。第二層では \mathbf{u} とその予測信号 $\boldsymbol{\theta}$ の比が入力され、入出力関数を経て第一層にフィードバックする。

これまでに述べた機構によって \mathbf{v} が複雑型細胞の応答特性を獲得したと仮定すると、一つの v_k はほぼ同じ位置の受容野を共有する u_j との結合が大きくなると考えられる。これは上で述べた第一層での抑制性の水平結合よりも局所的である。従って、モデル全体を考えると、学習の収束後は第一層のニューロン間にメキシカンハット状の相互作用が見られると考えられる。

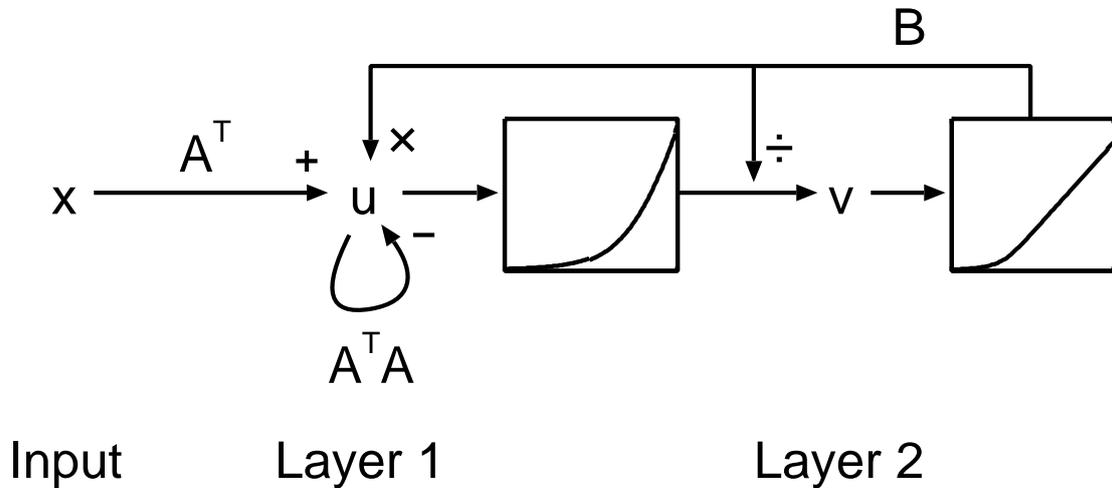


図 3.5: 事後分布の推定アルゴリズム.

3.6 まとめ

複雑型細胞は、様々な空間位相に応答する単純型細胞から興奮性の入力を受けることによって、位相不変な応答を実現していると考えられる。この中には、互いにほぼ逆位相の関係にある受容野の組み合わせも含まれている。このような単純型細胞の応答は明らかに負の相関を持つため、単純型細胞-複雑型細胞間の結合荷重の学習はSCのような線型な生成モデルによる学習則では説明できない。

本モデルは階層ベイズの枠組みによって、この複雑型細胞特性の獲得を統計的に説明することを目的としている。本章前半では、二階層の階層ベイズモデルを構築し、自然画像による学習を行った。学習後、第一層はGabor型の基底を獲得した。また第二層は、線型モデルでは獲得できない、複数のGabor型基底の組み合わせを獲得した。ただし、それは複雑型応答を構成するような、空間位相の組み合わせとはならなかった。この原因としては、第一層においてスパースネスが強すぎることにより、空間周波数と方位が等しく空間位相のみが異なる複数の基底が獲得されていないことが原因と考えられる。

後半では、非負制約を加えることによって、より生理学的に妥当なモデルを構築した。また、変分ベイズ法によってアルゴリズムを構築することで、単純型細胞のstatic nonlinearityに類似する非線型性など、V1でみられる特性が導出されることが示された。

第4章 生成モデルへのカーネル法の導入

4.1 目的

前章では生成モデルに求心性経路の非線型性を導出した。しかし、第一層のモデルは、応答特性は非線型であるが、本質的には基底ベクトルの線型和で入力表現するという線型のモデルである。モデル全体の非線型性は、第二層によるスケールの表現に依っている。また、第一層では加法的な表現、第二層では乗法的な表現となっていることから、各階層における表現の様式が大きく異なっているという問題がある。そこで本章では、生成モデルにより自然な形で非線形を導入する方法を検討する [86]。

Rao ら [51] の予測符号化モデルでは、下位の領野からの入力 \mathbf{x} を応答 \mathbf{y} の非線形変換 $f(A\mathbf{y})$ で予測する。これは遠心性結合を介した上位領野から下位領野への信号伝達を表している。このとき、最小化すべき目的関数は、以下のようになる：

$$E = \frac{1}{2} \|\mathbf{x} - f(A\mathbf{y})\|^2 + h(\mathbf{y}). \quad (4.1)$$

ここで $h(\mathbf{y})$ は、 \mathbf{y} が小さく、またはスパースになるための制約項である。視覚系はこの E を勾配に基づいて最小化しているとする、応答 \mathbf{y} の時間変化は

$$\begin{aligned} \Delta\mathbf{y} &\propto -\frac{\partial E}{\partial \mathbf{y}} \\ &= A^T \frac{df}{d\mathbf{y}}^T (\mathbf{x} - f(A\mathbf{y})) - \frac{dh}{d\mathbf{y}} \end{aligned} \quad (4.2)$$

で与えられる。Rao らは誤差信号 $\mathbf{x} - f(\mathbf{y})$ を符号化するニューロン集団を考え、その応答が求心性結合 A^T を介して伝搬すると考えた。また式 (4.2) の第一項を展開すれば、 \mathbf{x} の求心性入力と $f(\mathbf{y})$ による側方抑制を考えることもできる。どちらの場合も、この求心性入力は、少なくとも短期的にみれば \mathbf{x} に対して線型である。

一方、V1 ニューロンは expansive な非線型性を持つことが報告されている [2]。従って求心性入力は、expansive な関数 $g(u)$ を用いて $g(A^T \mathbf{x})$ と表現される方がよ

り妥当である. 式(4.2)を直接変更することで, 構成論的にこの非線型性の役割を調べることは可能である. しかし, この非線型性の役割を本質的に理解するためには, 式(4.2)が生成モデルから導出されたように, $g(A^T \mathbf{x})$ が自然に導出されるような枠組みを考える必要がある.

いま $f(\mathbf{y}) = A\mathbf{y}$ となる線型生成モデルを考える. このとき求心性の入力は $A^T \mathbf{x}$ となる. ここで $A^T \mathbf{x}$ の各成分をみると,

$$[A^T \mathbf{x}]_i = \mathbf{a}_i^T \mathbf{x} = \langle \mathbf{a}_i, \mathbf{x} \rangle \quad (4.3)$$

より, これらは入力 \mathbf{x} と A の基底ベクトル \mathbf{a}_i とのユークリッド空間における内積と解釈できる. そこで本章では, 生成モデルの枠組みにカーネル法を導入し, 非線型性の計算論的役割について議論する.

4.2 モデル

前節で述べた通り, 入力 $\mathbf{x} \in \mathbb{R}^m$ の求心性結合による線型な伝搬 $A^T \mathbf{x}$ は, \mathbb{R}^m 上の内積と考えることができる. ここから, \mathbf{x} と各基底 \mathbf{a}_j をある非線型写像 ϕ で特徴空間 \mathcal{F} に写し, それらの内積が

$$\langle \phi(\mathbf{a}_j), \phi(\mathbf{x}) \rangle_{\mathcal{F}} = f(\mathbf{a}_j^T \mathbf{x}) \quad (4.4)$$

となるように, カーネル法に基づくモデルを構築すればいいことがわかる. この節では, まず本モデルの構造を示し, 具体的なパラメータに依存しない一般的な性質について述べる.

4.2.1 非負性

本モデルでは生理学的妥当性のため, モデルパラメータを非負とする. 以下, m 次元の非負実数ベクトルの空間を \mathbb{R}_+^m と表す. いま, 平均が0となるように中心化された入力画像を \mathbf{x} とする. LGN では, on-center 型の受容野と off-center 型の受容野と rectifier によって, 周囲より明るい点と暗い点がそれぞれ別のニューロンによって符号化される. そこで視床からの入力 $\mathbf{x}^+ \in \mathbb{R}_+^m$ を

$$\mathbf{x}^+ = \frac{1}{2} \begin{pmatrix} |\mathbf{x}| + \mathbf{x} \\ |-\mathbf{x}| - \mathbf{x} \end{pmatrix} \quad (4.5)$$

とする. ここで $|\mathbf{x}|$ は \mathbf{x} の各成分について絶対値をとる演算を表す. この変換 $\mathbb{R}^m \ni \mathbf{x} \mapsto \mathbf{x}_+ \in \mathbb{R}_+^{2m}$ によって, ベクトルの次元は二倍になるが, $i = 1, \dots, m$ について x_i と x_{m+i} は以下のような冗長性を持っている.

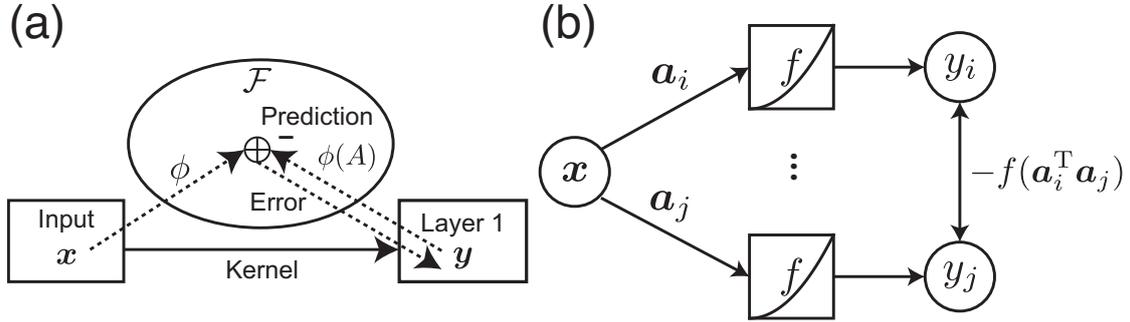


図 4.1: モデルの構造. (a) 特徴空間における予測符号化. (b) 求心性結合と側抑制.

1. $x_i^+ \neq 0$ ならば $x_{i+m}^+ = 0$.
2. $x_{i+m}^+ \neq 0$ ならば $x_i^+ = 0$.

$x_i = 0$ のとき $x_i^+ = x_{i+m}^+ = 0$ なので, これらの逆は成り立たない. \mathbf{x}^+ は逆変換

$$\mathbf{x} = \begin{pmatrix} I_m & -I_m \end{pmatrix} \mathbf{x}^+ \quad (4.6)$$

によって画像として再構成できる. ここで I_m は m 次の単位行列である. 式 (4.6) より, この変換が \mathbf{x} の基底 $\{\mathbf{e}_i\}_{i=1}^m \cup \{-\mathbf{e}_i\}_{i=1}^m$ (\mathbf{e}_i は \mathbb{R}^m 上の標準基底) による overcomplete な表現であるといえる.

V1 におけるニューロン集団の応答を $\mathbf{y} \in \mathbb{R}_+^n$ で表す. \mathbf{y} の各成分は V1 ニューロンの発火頻度をモデル化する. また, LGN から V1 ニューロン j への結合を $\mathbf{a}_j \in \mathbb{R}_+^{2m}$ とし, $A = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) \in \mathbb{R}_+^{2m \times n}$ とする. これは LGN から V1 への結合が全て興奮結合であることを表している.

以上のような定式化により, 全てのパラメータを非負ベクトルで表現した. 皮質モデルの計算としては, on-center 型と off-center 型の応答を区別する必要はなく, 単に LGN からの入力と考えてよい. そこで, 以下では m を偶数とし, 上で述べた \mathbf{x}^+ を改めて入力 \mathbf{x} として表す.

4.2.2 構造

カーネル法の原理に従い, 入力 \mathbf{x} と A の基底ベクトル $\mathbf{a}_j (j = 1, 2, \dots, n)$ を $\phi: \mathbb{R}^m \rightarrow \mathcal{F}$ によって特徴空間 \mathcal{F} に写像する. $\phi(\mathbf{x})$ は各成分 $\phi_i(\mathbf{x})$ を縦に並べた

ベクトルとする。また、 $\phi(\mathbf{x})^T \phi(\mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle_{\mathcal{F}}$ とする。

$$E(A, \mathbf{y}) = \frac{1}{2} \left\| \phi(\mathbf{x}) - \sum_{j=1}^n \phi(\mathbf{a}_j) y_j \right\|_{\mathcal{F}}^2 \quad (4.7)$$

$$= \frac{1}{2} \|\phi(\mathbf{x}) - \phi(A)\mathbf{y}\|_{\mathcal{F}}^2 \quad (4.8)$$

ここで、 $\phi(A) = (\phi(\mathbf{a}_1), \phi(\mathbf{a}_2), \dots, \phi(\mathbf{a}_n))$ とおいた。

$E(A, \mathbf{y})$ は \mathbf{y} に関して二次式なので、これを \mathbf{y} で偏微分すると

$$\frac{\partial E}{\partial \mathbf{y}} = \phi(A)^T (\phi(\mathbf{x}) - \phi(A)\mathbf{y}) \quad (4.9)$$

$$= \phi(A)^T \phi(\mathbf{x}) - \phi(A)^T \phi(A)\mathbf{y} \quad (4.10)$$

が得られる。

$\frac{\partial E}{\partial \mathbf{y}}$ の各成分は、

$$\frac{\partial E}{\partial y_j} = \langle \mathbf{a}_j, \mathbf{x} \rangle_{\mathcal{F}} - \sum_i \langle \mathbf{a}_i, \mathbf{a}_j \rangle_{\mathcal{F}} \quad (4.11)$$

と表せる。第一項は \mathbf{x} と \mathbf{a}_j の関数なので、ニューロン j への入力と考えることができる。

4.3 カーネル関数

1.5.1 で述べたように、画像のベクトル空間 \mathbb{R}^m から特徴空間 \mathcal{F} への写像 $\phi(\mathbf{x})$ が定まるためには、カーネル関数は正定値でなければならない。しかし、一般の対称性をもつ関数 $k: \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}$ は必ずしも正定値ではない。一方で後述するように、ニューロン応答特性をカーネル法と関連付けて考察するためには、不定値カーネルの振る舞いについても考える必要がある。この節では、ニューロン応答をモデル化するためにどのようなカーネル関数が適切であるかを考察する。またカーネル関数による特徴空間の性質と、特に不定値カーネルについてはその計算論的意義について議論する。

4.3.1 カーネル関数の条件

前節で述べたように、カーネル関数はニューロンへの求心性の入力を表している。式 (1.3) や式 (1.4) に示したように、神経回路の構成要素としてニューロンをみると、計算論的に重要である性質は、シナプス結合による重み付き和と、そ

れに対する非線型変換である。特に単純型細胞については、視野上の各点の輝度と発火頻度の入出力関係 [93] は式 (1.4) のような LN (linear-nonlinear) 型のモデルで説明できる。また、発火頻度は非負であり、線型フィルタ応答に対して単調増加する。

以上より、カーネル関数を

$$k(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}^T \mathbf{y}) \quad (4.12)$$

の形に制限することは妥当であると考えられる。ここで $f(u)$ は、 $u \geq 0$ で定義される C^∞ 級の非負関数である。また、 $f'(u) > 0$ とする。

4.3.2 カーネル関数と特徴空間の関係

このように表されるカーネル関数は、 \mathcal{X} 上の内積を含むことから、 $f(\cdot)$ の形状によらず対称である。しかし、正定値性については $f(\cdot)$ の形状に依存する。簡単のため、 $f(u)$ が C^∞ 級関数で、冪級数展開

$$f(u) = \sum_{k=0}^{\infty} f_k u^k \quad (4.13)$$

が存在すると仮定する。このとき $u = \mathbf{x}^T \mathbf{y}$ を代入すると、

$$f(\mathbf{x}^T \mathbf{y}) = \sum_{k=0}^{\infty} f_k \left(\sum_{i=1}^m x_i y_i \right)^k \quad (4.14)$$

となる。これを多項定理を用いて展開するために、指数ベクトル $\mathbf{p} \in \mathbb{N}^m$ (ここで $0 \in \mathbb{N}$) を

$$\mathbf{x}^{\mathbf{p}} = \prod_{i=1}^m x_i^{p_i}, \quad |\mathbf{p}| = \sum_{i=1}^m p_i \quad (4.15)$$

と定義する。また $|\mathbf{p}| = k$ となる \mathbf{p} の集合を P_k で表す。カーネル関数は多項定理により

$$\begin{aligned} f(\mathbf{x}^T \mathbf{y}) &= \sum_{k=0}^{\infty} f_k \sum_{\mathbf{p} \in P_k} \binom{k}{\mathbf{p}} \mathbf{x}^{\mathbf{p}} \mathbf{y}^{\mathbf{p}} \\ &= \sum_{\mathbf{p} \in \mathbb{N}^m} f_{|\mathbf{p}|} \binom{|\mathbf{p}|}{\mathbf{p}} \mathbf{x}^{\mathbf{p}} \mathbf{y}^{\mathbf{p}} \end{aligned} \quad (4.16)$$

と展開される。ここで

$$\binom{k}{\mathbf{p}} = \frac{k!}{\prod_{i=1}^m p_i!} \quad (4.17)$$

は多項係数を表す。従って、 m 個の非負整数の組 \mathbf{p} に対応して一つの基底関数

$$\phi_{\mathbf{p}}(\mathbf{x}) = \sqrt{f_{|\mathbf{p}|} \binom{|\mathbf{p}|}{\mathbf{p}}} \mathbf{x}^{\mathbf{p}} \quad (4.18)$$

が与えられ、その値を特徴ベクトル $\phi(\mathbf{x})$ の成分と考えることができる。 $\mathbf{p} \neq \mathbf{q}$ に対して $\mathbf{x}^{\mathbf{p}}$ と $\mathbf{x}^{\mathbf{q}}$ は \mathbf{x} の関数として一次独立であるため、 $f_k \neq 0$ となる全ての k に対する P_k の位数の総和が特徴空間の次元となる。これは高々加算無限である。そこで、以下では特徴ベクトルの成分の添字として自然数の代わりに係数ベクトルを用いる。

多項係数は常に正であるため、 $f_k > 0$ ならば対応する成分 (P_k に含まれる全ての \mathbf{p} に対して $\phi_{\mathbf{p}}$) は実数となる。また f_k が大きいほど対応する次元の広がりが大きくなる。一方 $f_k < 0$ の場合、対応する $\phi_{\mathbf{p}}$ は純虚数であり、実数では表せない。そこで、特徴ベクトルを実ベクトルとして表現するために、特徴空間の計量 (つまり内積の定義) を変更する。

$$K_+ = \{k \mid f_k > 0\}, \quad K_- = \{k \mid f_k < 0\} \quad (4.19)$$

とおくと、

$$f(u) = \sum_{k \in K_+} |f_k| u^k - \sum_{k \in K_-} |f_k| u^k \quad (4.20)$$

と表せるので、 $\mathbf{p}, \mathbf{q} \in \mathbb{N}^m$ に対応する基底ベクトルの内積を

$$\langle e_{\mathbf{p}}, e_{\mathbf{q}} \rangle_{\mathcal{H}} = \begin{cases} 1 & (\mathbf{p} = \mathbf{q} \wedge |\mathbf{p}| \in K_+) \\ -1 & (\mathbf{p} = \mathbf{q} \wedge |\mathbf{p}| \in K_-) \\ 0 & (\mathbf{p} \neq \mathbf{q}) \end{cases} \quad (4.21)$$

と定義すると、特徴空間を

$$\phi_{\mathbf{p}}(\mathbf{x}) = \sqrt{|f_{|\mathbf{p}|}| \binom{|\mathbf{p}|}{\mathbf{p}}} \mathbf{x}^{\mathbf{p}} \quad (4.22)$$

のように実数の成分を持つ空間として表現できる。

4.3.3 不定値カーネル

以上の定義により、 $f(u)$ の多項式展開が正の係数のみを持つとき特徴空間 \mathcal{H} は通常のヒルベルト空間となり、負の係数を含むとき \mathcal{H} の計量は不定値となる。例えば、シグモイドカーネルは式 (4.12) の形式を持つが、シグモイド関数の多項式

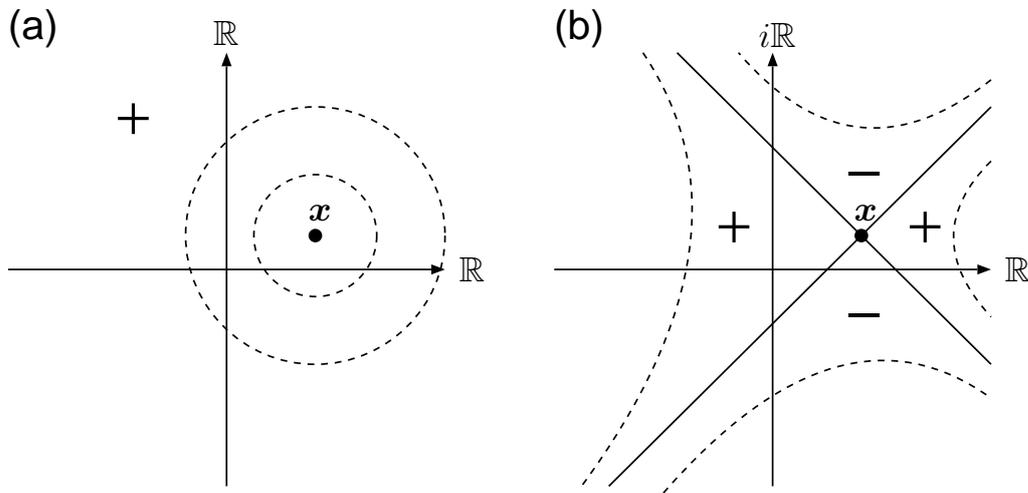


図 4.2: ユークリッド空間 \mathbb{R}^2 (a) と擬ユークリッド空間 $\mathbb{R}^{(1,1)}$ (b) における計量. 点 x と空間上の各点との距離の二乗 d^2 を等高線 (破線) で模式的に示す. $+$, $-$ は実線で区切られた各領域において d^2 がそれぞれ正, 負になることを表す. \mathbb{R} と $i\mathbb{R}$ はそれぞれ計量が正定値, 負定値となる部分空間を表す.

展開は負の係数を持つので, 対応する特徴空間は不定値となる. これはカーネル関数の条件として正定値性を採用すると, $f(u)$ として変曲点を持つような複雑な関数を用いることができないことを示す.

このように, カーネル関数に正定値性の条件を加えることは, ニューロン応答をモデル化する上で, 強い制約となる. 一方で, 4.3.1 節で述べたように, カーネル関数の非負性は妥当な条件であると考えられる. 非負性と不定値性は異なる条件であるが, 後述するように最適化問題の収束性に関して類似した働きを持つ. また SVM においては, 不定値カーネルの有用性が様々な研究によって示されている [94, 95, 96, 97, 98].

そこで, 本モデルではカーネル関数として不定値カーネルも考慮する. 不定値カーネルはヒルベルト空間上の内積として解釈することはできないが, 後述するように空間の定義を拡張することで, ある程度の幾何学的な解釈が可能となる. 例えば, SVM による分類問題においては Haasdonk[99] が幾何学的な考察を行っている. ここでは不定値カーネルに対応する特徴空間の特性について議論する.

簡単のため特徴空間 \mathcal{H} を有限次元とする. この場合, \mathcal{H} は擬ユークリッド空間として表せる [100, 101, 102].

4.3.4 カーネル関数の微分

本モデルでは，目的関数の勾配の導出や微分幾何学的な考察において，カーネル関数の偏微分を扱う．特に，偏微分によって対称性を失う場合や，特徴空間での二乗ノルムに相当する $k(\mathbf{x}, \mathbf{x})$ の x_i による偏微分などを扱うことになる．そこでまず，カーネル関数 k を C^∞ 級の関数と仮定し，偏微分の記法を定義するとともに， k の偏導関数の再生性について議論する．

$k(\mathbf{x}, \mathbf{y})$ は，再生核ヒルベルト空間 [103] \mathcal{H}_k 上における，実数値関数 $k(\mathbf{x}, \cdot), k(\mathbf{y}, \cdot)$ の内積として特徴付けられる． $k(\mathbf{x}, \cdot)$ 自体が \mathbf{x} の関数であるため， $\frac{\partial}{\partial x_i} k(\mathbf{x}, \cdot)$ などの偏微分を考えることができるが，これらは実数値関数であるので，やはり \mathcal{H}_k の元となる．そこで， $\frac{\partial}{\partial x_i} k(\mathbf{x}, \cdot)$ と， \mathbf{x} に独立な別の関数 $f(\cdot)$ の内積を取ると，内積の双線形性と微分の線形性より

$$\left\langle \frac{\partial}{\partial x_i} k(\mathbf{x}, \cdot), f(\cdot) \right\rangle = \frac{\partial}{\partial x_i} \langle k(\mathbf{x}, \cdot), f(\cdot) \rangle \quad (4.23)$$

となり，カーネル関数の再生性より

$$\left\langle \frac{\partial}{\partial x_i} k(\mathbf{x}, \cdot), f(\cdot) \right\rangle = \frac{\partial}{\partial x_i} f(\mathbf{x}) \quad (4.24)$$

が成り立つ．これは偏微分が再生性を保つことを示している．ここで $f(\cdot) = k(\mathbf{y}, \cdot)$ を考えると，

$$\left\langle \frac{\partial}{\partial x_i} k(\mathbf{x}, \cdot), k(\mathbf{y}, \cdot) \right\rangle_{\mathcal{H}_k} = \frac{\partial}{\partial x_i} k(\mathbf{x}, \mathbf{y}) \quad (4.25)$$

と書けるが， $\mathbf{x} = \mathbf{y}$ の場合に，右辺において $\frac{\partial}{\partial x_i}$ が左右どちらの引数に対する偏微分であるかが区別ができなくなる． $k(\mathbf{x}, \mathbf{x})$ は $k(\mathbf{x}, \cdot)$ の二乗ノルムを表し， $k(\mathbf{x}, \cdot)$ について非線形なので， $f(\cdot) = k(\mathbf{x}, \cdot)$ について式 (4.23) のような等式は成り立たない．従って， $k(\mathbf{x}, \cdot)$ の偏微分同士で内積を取る場合， $\frac{\partial}{\partial x_i}$ がどちらか一方の引数に関する偏微分であることを明記する必要がある．この場合

$$\left\langle \frac{\partial}{\partial x_i} k(\mathbf{x}, \cdot), k(\mathbf{x}, \cdot) \right\rangle_{\mathcal{H}_k} = \frac{\partial}{\partial x_i} k(\mathbf{x}, \mathbf{y}) \Big|_{\mathbf{y}=\mathbf{x}} \quad (4.26)$$

のように表現できるが，表記が煩雑になるので，次のように表現する．

$$k_{(i_1 \dots i_r)(j_1 \dots j_s)}(\mathbf{x}, \mathbf{y}) = \frac{\partial^{r+s}}{\partial \xi_{i_1} \dots \partial \xi_{i_r} \partial \zeta_{j_1} \dots \partial \zeta_{j_s}} k(\boldsymbol{\xi}, \boldsymbol{\zeta}) \Big|_{\boldsymbol{\xi}=\mathbf{x}, \boldsymbol{\zeta}=\mathbf{y}} \quad (4.27)$$

すなわち， k の添字のうち，左の括弧は k の左の．ここで， $k_{()} = k$ とする．また，カーネル関数の対称性より

$$k_{(i_1 \dots i_r)(j_1 \dots j_s)}(\mathbf{x}, \mathbf{y}) = k_{(j_1 \dots j_s)(i_1 \dots i_r)}(\mathbf{y}, \mathbf{x}) \quad (4.28)$$

である。従って式 (4.26) の場合, $k_{(i)()}(\mathbf{x}, \mathbf{x})$ または $k_{()}(i)(\mathbf{x}, \mathbf{x})$ と書ける。

一方, 内積を求めた後の偏微分についても考える必要がある。内積の左右のオペランドが独立であれば

$$\frac{\partial}{\partial x_l} k_{(i_1, \dots, i_r)(j_1, \dots, j_s)}(\mathbf{x}, \mathbf{y}) = k_{(i_1, \dots, i_r, l)(j_1, \dots, j_s)}(\mathbf{x}, \mathbf{y}) \quad (4.29)$$

と書けるので, そうでない場合について考える。この場合, 左右のオペランドは偏微分の仕方によって異なる関数となるが, どちらも \mathbf{x} の関数である。また, ある i に対して x_i に関する偏微分を考えるので, 他の変数を固定して一変数の関数とみなせる。そこで, 一般の双線形で対称な内積 $\langle \cdot, \cdot \rangle_{\mathcal{X}}$ について $\frac{d}{dx} \langle f(x), g(x) \rangle_{\mathcal{X}}$ を考える。微分の定義より

$$\frac{d}{dx} \langle f(x), g(x) \rangle_{\mathcal{X}} = \lim_{h \rightarrow 0} \frac{\langle f(x+h), g(x+h) \rangle_{\mathcal{X}} - \langle f(x), g(x) \rangle_{\mathcal{X}}}{h}. \quad (4.30)$$

ここで $f(x+h), g(x+h)$ を x の周りでテイラー展開すると,

$$\begin{aligned} \frac{d}{dx} \langle f(x), g(x) \rangle_{\mathcal{X}} &= \lim_{h \rightarrow 0} \frac{\langle f(x) + f'(x)h, g(x) + g'(x)h \rangle_{\mathcal{X}} - \langle f(x), g(x) \rangle_{\mathcal{X}} + o(h^2)}{h} \\ &= \lim_{h \rightarrow 0} \frac{h \langle f(x), g'(x) \rangle_{\mathcal{X}} + h \langle f'(x), g(x) \rangle_{\mathcal{X}} + o(h^2)}{h} \\ &= \lim_{h \rightarrow 0} \langle f(x), g'(x) \rangle_{\mathcal{X}} + \lim_{h \rightarrow 0} \langle f'(x), g(x) \rangle_{\mathcal{X}} + \lim_{h \rightarrow 0} o(h) \\ &= \langle f(x), g'(x) \rangle_{\mathcal{X}} + \langle f'(x), g(x) \rangle_{\mathcal{X}} \end{aligned} \quad (4.31)$$

が得られるので, x に依存する二変数の内積は, 微分に関して関数の積のように振る舞うことがわかる。これを用いて

$$\frac{\partial}{\partial x_l} k_{(i_1, \dots, i_r)(j_1, \dots, j_s)}(\mathbf{x}, \mathbf{x}) = (k_{(i_1, \dots, i_r, l)(j_1, \dots, j_s)} + k_{(i_1, \dots, i_r)(j_1, \dots, j_s, l)})(\mathbf{x}, \mathbf{x}) \quad (4.32)$$

と表せる。

4.4 最適化問題の収束性

目的関数 $E(A, \mathbf{y})$ は \mathbf{y} に関して二次式であり, 最適化問題の凸性はグラム行列 $G_A = \{k(\mathbf{a}_i, \mathbf{a}_j)\}_{ij}$ の正定値性と同値である。従ってカーネル関数が正定値であれば, $E(A, \mathbf{y})$ が \mathbf{y} に関して凸関数となり, \mathbf{y} の最適解が常に存在する。一方カーネル関数が不定値のとき, グラム行列は必ずしも正定値にはならない。この場合, 最適化問題の収束性は自明ではない。これが収束しない場合, A, \mathbf{y} について何らか

の制約条件を設ける必要がある。そこで、この節では最適化問題の収束性について議論する。

まず \mathbf{y} については、不定値行列 $Q \geq 0$ を用いて以下の問題に置き換えられる。

$$\min_{\mathbf{y}} F_1(\mathbf{y}) = \mathbf{y}^T Q \mathbf{y} + \mathbf{b}^T \mathbf{y} \quad (4.33)$$

$$\text{s.t.} \quad \mathbf{y} \geq \mathbf{0} \quad (4.34)$$

ここで Q は零行列 O ではないとする。もし $\mathbf{y} \geq \mathbf{0}$ の範囲で常に $F(\mathbf{y}) \geq 0$ であるならば、 $F(\mathbf{y})$ の最小値が存在するので、最適解も存在する。そこで、 $S = \{\mathbf{y} \mid \mathbf{y} \geq \mathbf{0}, F(\mathbf{y}) < 0\}$ が空集合でない場合を考える。このとき最適解は S に含まれるので、 S が有界であれば解は収束する。 $F(\mathbf{0}) = 0$ なので $\mathbf{0} \notin S$ である。ここで、 $F(\mathbf{y})$ は \mathbf{b} と \mathbf{y} の方向余弦 $\cos \theta_{b\mathbf{y}}$ を用いて

$$F_1(\mathbf{y}) = \mathbf{y}^T Q \mathbf{y} + \|\mathbf{y}\| \|\mathbf{b}\| \cos \theta_{b\mathbf{y}} \quad (4.35)$$

と書ける。 $\mathbf{y} > 0$ の場合、第一項は常に正值で、その絶対値は $\|\mathbf{y}\|^2$ に比例する。一方、第二項の符号は \mathbf{y} の方向によって異なるが、その絶対値は $\|\mathbf{y}\|$ に比例する。従って、 $\|\mathbf{y}\|$ を十分大きくすることで、第一項の絶対値を第二項の絶対値よりも大きくすれば、 $F_1(\mathbf{y}) \leq 0$ 、つまり $\mathbf{y} \notin S$ とすることができる。よって S は有界であり、 $F_1(\mathbf{y})$ を最小にする \mathbf{y} が存在する。

次に、行列 A について述べる。非線型写像 ϕ のため、 $E(A, \mathbf{y})$ は A の各成分について二次式とはならない。そこでまず $\phi(A) = (\phi(\mathbf{a}_1), \dots, \phi(\mathbf{a}_n))$ について考えると、 $E(A, \mathbf{y})$ は $\phi(A)$ の各成分について二次式となる。いま、特徴空間が有限次元であるとして、式 (4.21) を対角成分とする対角行列を M で表す。このとき、目的関数は

$$E(A, \mathbf{y}) = (\phi(\mathbf{x}) - (\mathbf{y}^T \otimes I) \text{vec}(\phi(A)))^T M (\phi(\mathbf{x}) - (\mathbf{y}^T \otimes I) \text{vec}(\phi(A))) \quad (4.36)$$

と書ける。ここで vec は行列の列ベクトルを縦に列挙して列ベクトルを作る演算子である。つまり $\text{vec}(\phi(A))$ は、 $\phi(\mathbf{a}_i)$ を列ベクトルとして表して縦に並べた列ベクトルである。また \otimes はクロネッカー積を表す。式 (4.36) を展開して定数項を除くと

$$F_2(A) = \text{vec}(\phi(A))^T (\mathbf{y} \mathbf{y}^T \otimes M) \text{vec}(\phi(A)) - 2(\mathbf{y} \otimes M \phi(\mathbf{x}))^T \text{vec}(\phi(A)) \quad (4.37)$$

となり、 $F_1(\mathbf{y})$ と同様に $F_2(A)$ の最小値の存在を示すことができる。すなわち、第一項は \mathbf{y} の二次形式 $\mathbf{y}^T G_A \mathbf{y}$ を $\text{vec}(\phi(A))$ の二次形式に書き換えたものなので、 $\mathbf{y} \neq \mathbf{0}$ ならば常に正值をとる。また、第一項と第二項の絶対値はそれぞれ $\|\text{vec}(\phi(A))\|^2$

と $\|\text{vec}(\phi(A))\|$ に比例する。従って $F_1(\mathbf{y})$ と同様に、 $\|\text{vec}(\phi(A))\| < \infty$ について $F_2(A)$ の最小値が存在する。以上より、

$$\|\text{vec}(\phi(A))\|^2 = \sum_{i=1}^n \phi(\mathbf{a}_i)^T \phi(\mathbf{a}_i) = \sum_{i=1}^n k(\mathbf{a}_i, \mathbf{a}_i) < \infty \quad (4.38)$$

であるので、任意の i に対して $k(\mathbf{a}_i, \mathbf{a}_i)$ が発散しないことが示せた。これは \mathbf{a}_i が発散しないことと同値ではない。例えば、 $k(\mathbf{x}, \mathbf{x})$ が最大値を持たないがある実数 $r > 0$ を上限に持ち、 $\|\mathbf{x}\| \rightarrow \infty$ に対して r に漸近するような場合、 $k(\mathbf{x}, \mathbf{x})$ を r 以上の任意の実数に近づけるために、 $\|\mathbf{x}\|$ を任意に大きくすることができる。このとき $\sum_{i=1}^n k(\mathbf{a}_i, \mathbf{a}_i) < nr$ なので、 $F_2(A)$ を最小にする $\text{vec}(\phi(A))$ が nr を超える場合に A は常に Frobenius ノルムの意味で発散することになる。逆に、 $\|\mathbf{x}\| \rightarrow \infty$ のとき発散するようなカーネル関数の場合、任意の $r > 0$ について $k(\mathbf{a}_i, \mathbf{a}_i) = r$ を満たす \mathbf{a}_i が存在するので、 A は収束する。例えば、シグモイドカーネルは前者、 $f(u) = x^a, a \in \mathbb{N}$ の場合は後者に当たる。

以上の結果をまとめると、正定値性の代わりに非負性を条件とするカーネル関数について、次のことがいえる。 A を固定したとき、 $E(A, \mathbf{y})$ の最小化問題は収束する。一方、 \mathbf{y} を固定した A に関する $E(A, \mathbf{y})$ の最小化問題は、カーネル関数の形状によっては収束しない。従って、最小化問題の収束性は 4.2 節で述べた条件により保証される。

4.5 数値実験

以上で述べた通り、入力 \mathbf{x} と基底ベクトル \mathbf{a}_j を特徴空間に写像することで、求心性経路に非線型性を導入することができる。また、カーネル関数は正定値であることが幾何学的な解釈としては望ましいが、基底ベクトルの制約条件を加えることで収束性を保証できた。ただし、最も単純で生理学的妥当性が高いと思われるカーネル関数は、expansive な非線型性を持つ LN 型のものであると考えられる。そこでこの節では、本モデルの特性を確認するために自然画像の学習を行い、学習後のモデル特性について考察する。

4.5.1 方法

カーネル関数には $f(u) = u^\alpha$ の LN 型のものを用いた。 $\alpha > 1$ のとき、このカーネルは expansive な非線型性を持つ。入力画像には、自然画像から抽出した 16×16 ピクセルのパッチを用いた。これらは 4.2.1 節で述べたように、中心化された後正

負の信号に分けられた。従って入力ベクトルは512次元である。基底画像の数は400とした。以上の条件の下で \mathbf{y} と A を目的関数の勾配に基づいて更新した。

4.5.2 基底

まず、学習後に A の基底ベクトルとして獲得された画像を図4.3に示す。これらは4.2.1節で述べた逆変換により画像として再構成されている。 α が小さいとき、点状の基底画像が獲得された。また、 α が大きくなるに従って基底画像は方位選択性を持つようになった。

4.5.3 スパースネス

次にニューロン応答 \mathbf{y} の分布について述べる。 \mathbf{y} が V1 のポピュレーションを再現するならば、その分布はスパースになると考えられる。これを確かめるため、スパースネスを以下のように定義する。

まず、

$$a = \frac{\left(\frac{1}{n} \sum_i y_i\right)^2}{\frac{1}{n} \sum_i y_i^2} \quad (4.39)$$

とおく。これは \mathbf{y} と $\mathbf{1} = (1, \dots, 1)^T$ の方向余弦の二乗であるため、 \mathbf{y} の全ての成分が同じ値を取れば $a = 1$ となる。また、 \mathbf{y} の非負性より \mathbf{y} と $\mathbf{1}$ が直交することはないので $a > 0$ であり、 a が0に近づくためには \mathbf{y} の多くの成分が0に近くなければならない。従って、

$$\text{sparseness} = 1 - a \quad (4.40)$$

と定義すると、これは非負ベクトルのスパースネスをよく表しているといえる。

このスパースネスを図4.4に示す。横軸は α 、縦軸はスパースネスを表す。この図より、応答のスパースネスは α の増加に伴って単調に増加していることがわかる。

4.5.4 方位選択性

上で述べた通り、 α の増加に伴って基底画像が Gabor 型に近づいていることがわかる。そこで、実際にニューロンの方位選択性が増加することを定量的に確かめるため、以下のように生理実験で用いられている運動方向選択性の指標を用いて検証した。

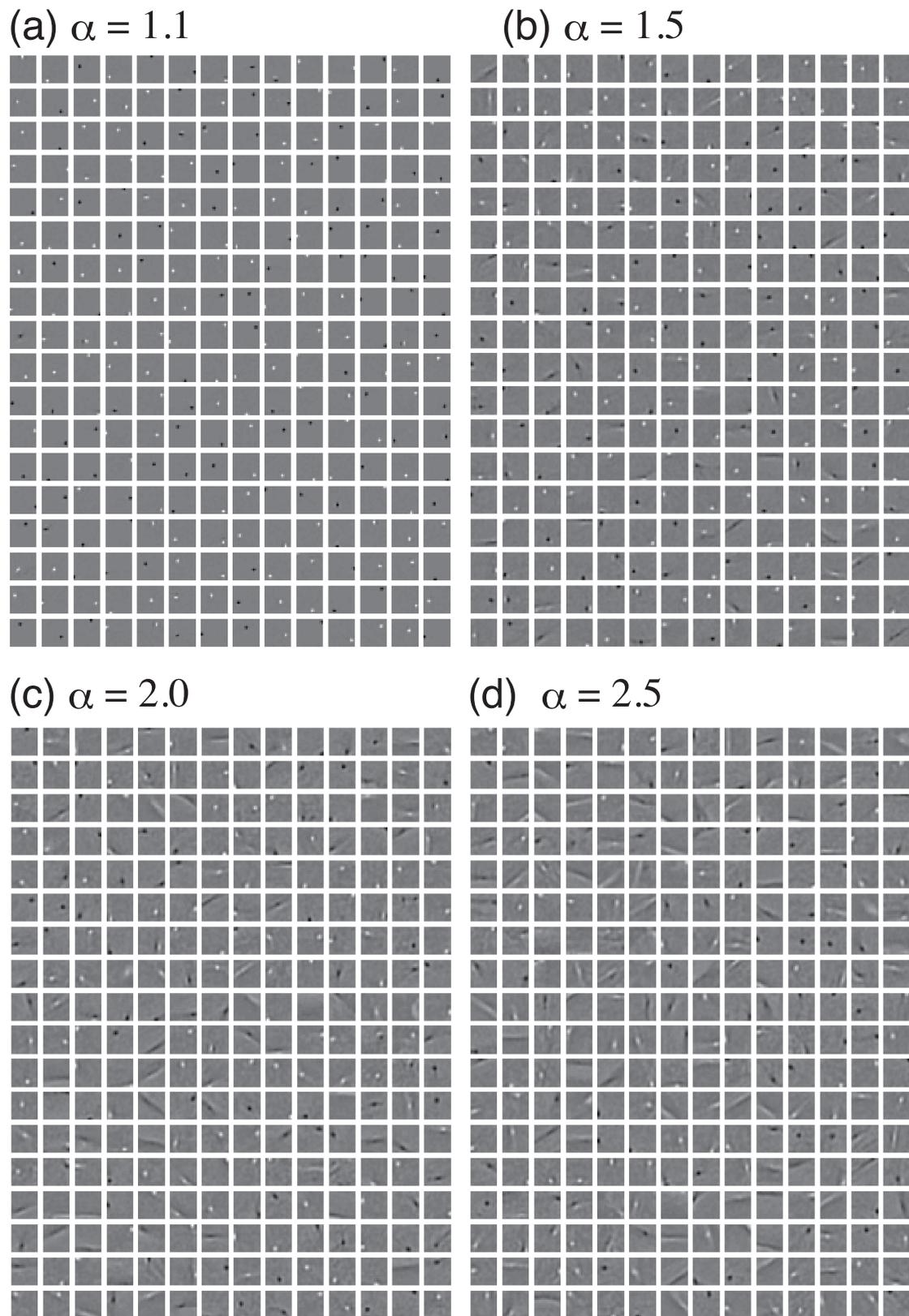


図 4.3: 学習後に獲得された基底画像. (a) $\alpha = 1.1$. (b) $\alpha = 1.5$. (c) $\alpha = 2.0$. (d) $\alpha = 2.5$.

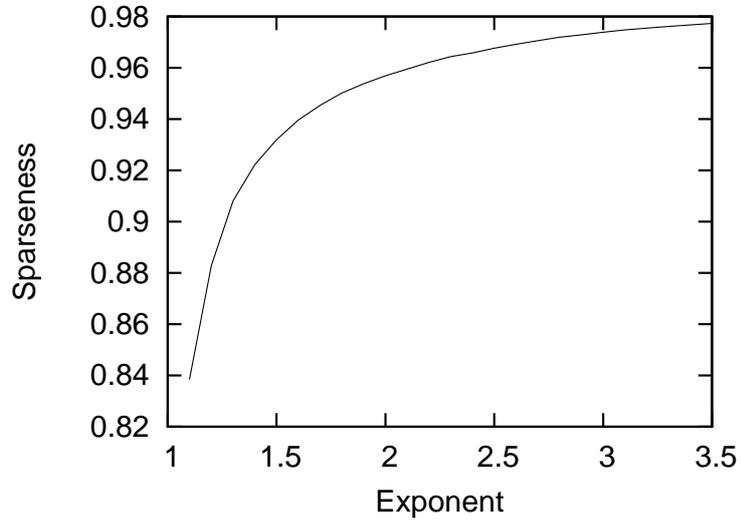


図 4.4: モデルニューロンのスパースネス.

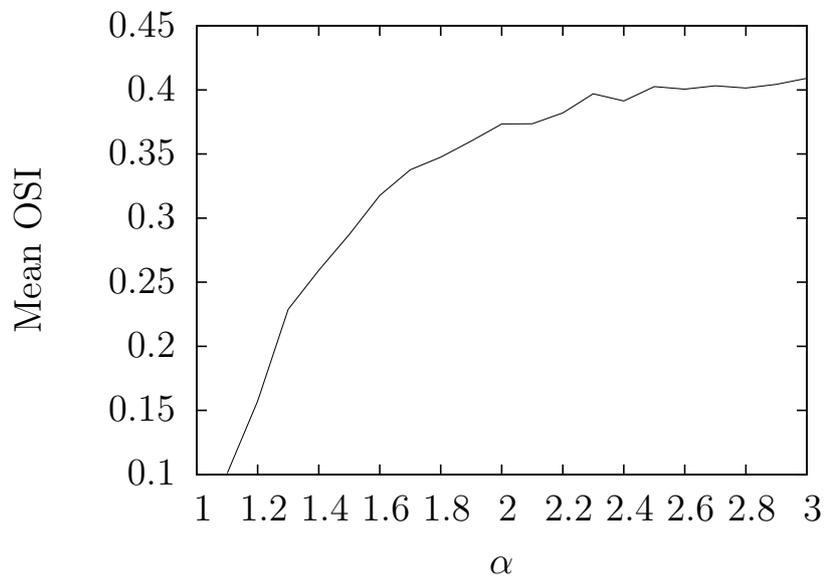


図 4.5: OSI の平均.

θ 方向の運動に対する応答を $R(\theta)$ とすると, $R(\theta)$ は周期 $T = 2\pi$ の周期関数とみることができる. そこで, $R(\theta)$ のフーリエ級数展開

$$R(\theta) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos k\theta + b_k \sin k\theta) \quad (4.41)$$

を考える. また, 各成分のパワーを

$$A_k = \sqrt{a_k^2 + b_k^2} \quad (k = 1, \dots, \infty), \quad A_0 = |a_0| \quad (4.42)$$

とする. もし対象のニューロンが特定の運動方向に選択性を持つならば, $R(\theta)$ は少なくとも周期 T の成分 A_1 が比較的大きいと考えられる. 逆に, このニューロンが運動方向の選択性を全く持たないのであれば, $R(\theta)$ はほぼ直流成分 A_0 のみを含んでいるといえる. 一方, 運動方向には選択性を持たず, 縞模様の方位に選択性を持つ場合, θ 方向の運動と $\theta + \pi$ 方向の運動に対する応答がほぼ等しく, $R(\theta)$ は周期 $T/2$ の成分 A_2 を含んでいると考えられる. そこで, 方位選択性の指標として OSI (Orientation Selectivity Index) を次のように定義する [104].

$$\text{OSI} = \frac{A_2}{A_0 + A_2}, \quad (4.43)$$

$$A_0 = \frac{1}{\pi} \left| \int_0^{2\pi} R(\theta) d\theta \right|, \quad (4.44)$$

$$A_2 = \frac{1}{\pi} \sqrt{\left(\int_0^{2\pi} R(\theta) \cos 2\theta d\theta \right)^2 + \left(\int_0^{2\pi} R(\theta) \sin 2\theta d\theta \right)^2} \quad (4.45)$$

OSI は 0 から 1 の値を取る. 対象のニューロンが方位選択性を全く持たない場合は OSI が 0 の値を取り, 方位選択性が強くなるに従って OSI は 1 に近づく. 本モデルでは $R(\theta) \geq 0$ を考えているので, $R(\theta)$ は $A_0 \geq 2A_2$ の直流成分を常に含む. 従って OSI は 1/3 以下となる.

全てのニューロンの OSI を平均したものを図 4.5 に示す. この図より, 方位選択性が α に伴って増加していることがわかる.

4.6 考察

Expansive な非線型性を持つ LN 型のカーネル関数を用いることで, 応答 \mathbf{y} はスパースに分布した. これは, 求心性の入力がこのような非線型変換を受けることで, 小さな入力信号はほぼ 0 付近に抑えられ, 大きな入力信号だけが強調されたためと考えられる.

また、同様のカーネル関数のもとでの自然画像の学習後、本モデルは Gabor 型の基底画像を獲得した。これについては、上記の理由によりポピュレーションの活動がスパースになっているため、結果として少ない基底で入力画像を表現しなければならず、スパースコーディングに類似する結果が得られたと考えられる。

4.6.1 ニューロン間の相互作用

カーネル関数による非線型性は、ニューロンへの求心性入力だけでなく、ニューロン間の側方結合にも影響する。この側方結合は、周辺抑制や方位交差抑制に関わっていると考えられる。そして、それはカーネル関数の形状に依存する。そこで、ここでは2ニューロンの場合について、側方結合を介したニューロン間の相互作用がカーネル関数の形状によってどのように変わるかを解析的に考察する。

いま、2ニューロンについて考えるので、 $\mathbf{y} = (y_1, y_2)^T$ とする。これらの受容野はそれぞれ $\mathbf{a}_1, \mathbf{a}_2$ である。多くの生理実験では、ニューロンの最適刺激を呈示しながら、それ以外の刺激を重畳することで周辺抑制や方位交差抑制の分析を行っている。そこで、入力を $\mathbf{x} = \alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2$ のように、ニューロン1の最適刺激である \mathbf{a}_1 とニューロン2の最適刺激である \mathbf{a}_2 との線型和とする。また、カーネル関数は $k(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}^T \mathbf{y})$ とし、

$$A^T A = \begin{pmatrix} 1 & c \\ c & 1 \end{pmatrix}, \quad (4.46)$$

つまり $\mathbf{a}_1, \mathbf{a}_2$ は正規ベクトルで方向余弦を c とする。基底ベクトルの非負性から $0 \leq c \leq 1$ である。

以上のように定義すると、目的関数は

$$\begin{aligned} E_1 &= \|\phi(\mathbf{x}) - \phi(\mathbf{a}_1)y_1 - \phi(\mathbf{a}_2)y_2\|_{\mathcal{H}} \\ &= f(1)y_1^2 + 2f(c)y_1y_2 + f(1)y_2^2 \end{aligned} \quad (4.47)$$

$$-2f(\alpha_1 + \alpha_2c)y_1 - 2f(\alpha_1c + \alpha_2) + f(1) \quad (4.48)$$

となる。これを平方完成して \mathbf{y} に依らない項を除いたものを

$$E_2 = (\mathbf{y} - \boldsymbol{\mu})^T A (\mathbf{y} - \boldsymbol{\mu}) \quad (4.49)$$

とする。ここで

$$A = \begin{pmatrix} f(1) & f(c) \\ f(c) & f(1) \end{pmatrix} \quad (4.50)$$

$$\boldsymbol{\mu} = \begin{pmatrix} f(1)f(\alpha_1 + \alpha_2c) - f(c)f(\alpha_1c + \alpha_2) \\ f(1)f(\alpha_1c + \alpha_2) - f(c)f(\alpha_1 + \alpha_2c) \end{pmatrix} \Big/ (f(1)^2 - f(c)^2) \quad (4.51)$$

である. E_2 を最小にする $y_1, y_2 \geq 0$ が応答となる. ここで $\boldsymbol{\mu} = (\mu_1, \mu_2)$ の成分の符号によって以下の背反な 4 通りの場合を考える.

1. $\mu_1 \geq 0 \wedge \mu_2 \geq 0$
2. $\mu_1 < 0 \wedge \mu_2 \geq 0$
3. $\mu_1 < 0 \wedge \mu_2 < 0$
4. $\mu_1 \geq 0 \wedge \mu_2 < 0$

1. の場合, 極値 $\boldsymbol{\mu}$ が非負条件を満たしているので, $\boldsymbol{y} = \boldsymbol{m}\boldsymbol{u}$ となる. 3. は, $0 \leq \frac{f(c)}{f(1)} \leq 1$ より $f(\alpha_1 + \alpha_2 c) < f(\alpha_1 c + \alpha_2)$ かつ $f(\alpha_1 c + \alpha_2) < f(\alpha_1 + \alpha_2 c)$ が必要条件となるが, これは成り立たない. 2. の場合は, 解の非負条件より $y_1 = 0$ となり,

$$E_2 = f(1)y_2^2 - 2f(\alpha_1 c + \alpha_2)y_2 = f(1) \left(y_2 - \frac{f(\alpha_1 c + \alpha_2)}{f(1)} \right)^2 + \text{const.} \quad (4.52)$$

より, $\boldsymbol{y} = (0, f(\alpha_1 c + \alpha_2))^T$ が解となる. 4. の場合も 2. と同様に, $\boldsymbol{y} = (f(\alpha_1 + \alpha_2), 0)^T$ となる. 以上をまとめると,

$$\boldsymbol{y} = \begin{cases} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} & (\mu_1 \geq 0 \wedge \mu_2 \geq 0) \\ \begin{pmatrix} f(\alpha_1 + \alpha_2 c)/f(1) \\ 0 \end{pmatrix} & (\mu_1 \geq 0 \wedge \mu_2 < 0) \\ \begin{pmatrix} 0 \\ f(\alpha_1 c + \alpha_2)/f(1) \end{pmatrix} & (\mu_1 < 0 \wedge \mu_2 \geq 0) \end{cases} \quad (4.53)$$

と表せる. これは α_1, α_2, c の区分関数である. 以下では $\alpha_1 = 0.5$ を固定し, y_1 を α_2 と c の関数とみて y_1 の最適刺激である \boldsymbol{a}_1 に y_2 の最適刺激である \boldsymbol{a}_2 を重畳するときの y_1 の応答について考察する.

線型モデル ($f(t) = t$) の場合, c に依らず $y_1 = \alpha_1$ となることから, \boldsymbol{a}_2 の重畳は y_1 の最終的な値に影響しない. $f(t)$ が expansive な非線型性を持つ場合, 図 4.6(a) のように, \boldsymbol{a}_2 の重畳によって y_1 の応答は増強されることになる. これは, 側方抑制の増加を求心性の入力の増加が上回っているためである. 逆に次第に緩やかになっていくような非線型性の場合, 図 4.6(b) のように, \boldsymbol{a}_2 の重畳は抑制的に働く.

以上の結果は, expansive な非線型性が側方抑制の効果を相対的に弱めることを示している. V1 でみられる方位交差抑制は LGN における応答の飽和によって説明できることが報告されている [105, 106] が, 皮質における expansive な非線型性がこの効果に拮抗していることが示唆される.

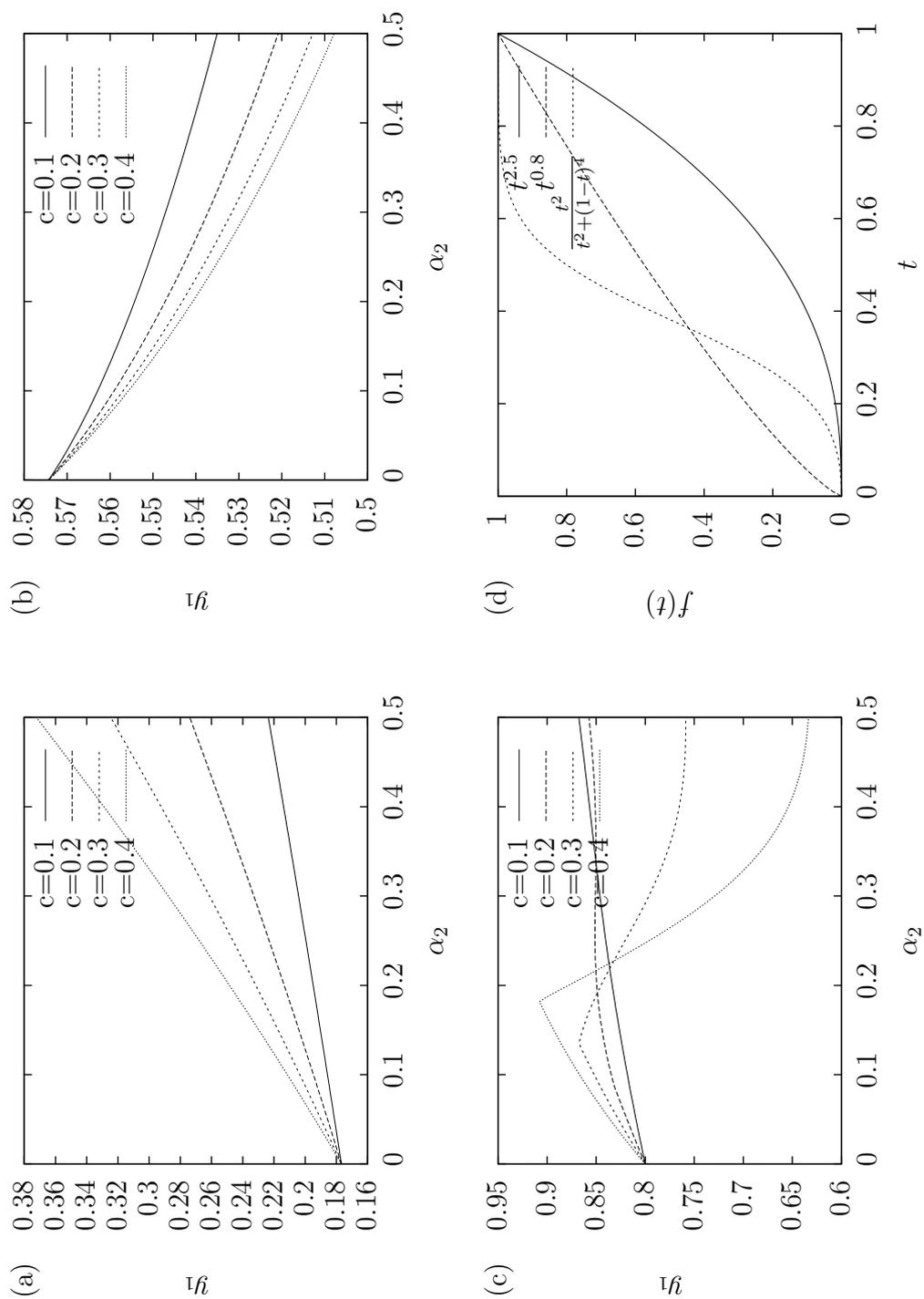


図 4.6: カーネル関数の非線型性とニューロン間の相互作用との関係. (a) $f(t) = t^{2.5}$ の場合. (b) $f(t) = t^{0.8}$ の場合. (c) $f(t) = \frac{t^2}{t^2 + (1-t)^4}$ の場合. (d) (a)-(c) で用いた $f(t)$ の形状.

4.6.2 基底ベクトルの正規化

これまでの議論では、基底ベクトル \mathbf{a}_j のノルムを1に固定してきた。1.3.3節で述べた通り、線型の生成モデルでは線型変換 A に制約を課す必要がある。このようなモデルでは、

$$\mathbf{x}_t \approx A\mathbf{y}_t, t = 1, \dots, T \quad (4.54)$$

を満たすような A と \mathbf{y}_t のうち、各 \mathbf{y}_t がある尺度において小さくなるようなものを選択される。ここで任意の正則な対角行列 D について $A' = AD^{-1}$, $\mathbf{y}'_t = D\mathbf{y}_t$ とおけば $A\mathbf{y}_t = A'\mathbf{y}'_t$ となるので、 A の各列を定数倍することで \mathbf{y} の各成分は任意に小さくできる。従って、 \mathbf{y} の各成分を小さくするためには対応する A の列を大きくすればよく、正規化はほとんど意味を成さない。このため A に対しても何らかの制約が設けられる。その一つとして考えられるものが A の列ベクトル \mathbf{a}_j の正規化である。

逆に言えば、 \mathbf{a}_j を正規化する代わりに y_j を $\|\mathbf{a}_j\|$ 倍することで同様の表現が得られるため、 y_j の事前分布が特異な形状でない限り、一般性を失うことなく \mathbf{a}_j を正規化することができると考えられる。このように最適化問題が \mathbf{a}_j の大きさについて冗長であるのは、目的関数が \mathbf{a}_j について線型であることによる。一方、本モデルでは、 \mathbf{a}_j は非線型関数 ϕ によって特徴空間に写像されるため、 \mathbf{a}_j を正規化することによって一般性を失う可能性があることに注意しなければならない。 \mathbf{a}_j をどのように制約すればよいかは \mathbf{x} の分布に依存すると考えられる。LN型のカーネル関数を用いる場合、 \mathbf{x} や \mathbf{a}_j のユークリッド内積が意味を持つ。本章では、 $\mathbf{x}^T \mathbf{a}_j$ が1以下になるように、 $\|\mathbf{a}_j\| = 1$ とした。

4.6.3 多層化

Raoらの予測符号化モデルと同様に、本モデルも多層化が可能である。ここでは二階層のモデルについて考える。

いま第一層の入力を $\mathbf{x} \in \mathcal{X}$ 、基底のセットを $\{\mathbf{a}_j \in \mathcal{X}\}_{j=1}^J$ とする。同様に第二層の基底のセットを $\{\mathbf{b}_k \in \mathcal{Y}\}_{k=1}^K$ で表す。また、第一層と第二層の出力をそれぞれ $\mathbf{y} \in \mathcal{Y}$, $\mathbf{z} \in \mathcal{Z}$ とする。 \mathcal{X} は任意の空間だが \mathcal{Y} , \mathcal{Z} はそれぞれ J 次元, K 次元のベクトル空間でなければならない。第一層では入力と基底が写像 $\phi: \mathcal{X} \rightarrow \mathcal{F}$ によって特徴空間 \mathcal{F} に写される。同様に第二層では $\psi: \mathcal{Y} \rightarrow \mathcal{G}$ によって写像される。第一層の出力 $\mathbf{y} \in \mathcal{Y}$ は第二層の入力となり、第二層では \mathbf{y} が写像 ψ によって特徴空間 \mathcal{G} に写される。

このモデルの目的関数は

$$E = \frac{1}{2} \left\| \phi(\mathbf{x}) - \sum_{j=1}^J y_j \phi(\mathbf{a}_j) \right\|_{\mathcal{F}}^2 + \frac{1}{2} \left\| \psi(\mathbf{y}) - \sum_{k=1}^K z_k \psi(\mathbf{b}_k) \right\|_{\mathcal{G}}^2 \quad (4.55)$$

となる。一階層のモデルと大きく異なる点は、第二層から第一層へのトップダウンの信号が導出されることである。\$E\$の第二項を\$E_2\$とすると、第二層から第一層のニューロン\$j\$へのフィードバックは、

$$-\frac{\partial E_2}{\partial y_j} = -\Psi_{(j)()}(\mathbf{y}, \mathbf{y}) + \sum_k z_k \Psi_{(j)()}(\mathbf{y}, \mathbf{b}_k) \quad (4.56)$$

となる。ここで\$\Psi\$は\$\psi\$に対応するカーネル関数である。さらに、LN型のカーネル関数\$\Psi(\mathbf{y}_1, \mathbf{y}_2) = g(\mathbf{y}_1^T \mathbf{y}_2)\$の場合は

$$-\frac{\partial E_2}{\partial \mathbf{y}} = -g'(\mathbf{y}^T \mathbf{y}) \mathbf{y} + \sum_k z_k g'(\mathbf{y}^T \mathbf{b}_k) \mathbf{b}_k \quad (4.57)$$

となる。第一項は\$\psi(\mathbf{y})\$を\$\mathcal{G}\$上のノルムの意味で小さくする効果を持つ。従って、第二層の追加によって\$\mathbf{y}\$の正則化項が自然に導入されたことになる。また第二項は、第二層の応答\$z\$が遠心性結合\$B^T = (\mathbf{b}_1, \dots, \mathbf{b}_K)^T\$を介して伝搬することを示している。

ここでは二階層のモデルについて述べたが、三階層以上のモデルについても同様に構築することができる。次節で述べるが、特徴空間へ写像することによって、ニューロンが表現する画像の位相構造を変化させることができる。多層化することにより、さらに複雑な特徴を表現するような特徴空間の構成が可能になると期待できる。

4.7 幾何学的考察

カーネル関数によって決定される特徴空間\$\mathcal{F}\$は一般に無限次元であるため、\$\mathcal{F}\$上でのモデルの振る舞いを分析することは難しい。そこで、ここでは本モデルの直感的な解釈を与えるため、幾何学的な考察を行う。

まず図4.7のような低次元の特徴空間の例を与える。一次元のベクトル\$x\$を特徴ベクトル\$\phi(\mathbf{x}) = (x, x^2)^T\$に写像することを考える。カーネル関数は\$k(x, y) = xy + x^2y^2\$である。この場合、特徴ベクトルは二次元となる。しかし、\$\phi_2 = \phi_1^2\$の関係が常に成り立つので、\$x\$がどのような値をとっても、\$\phi(\mathbf{x})\$は一つのパラメータで表せる曲線\$c(t) = (t, t^2)\$（つまり放物線）上に常に存在することになる。これ

は一次元多様体 \mathcal{M} とみなせる. 一方, $\phi(x)$ と $\phi(y)$ の距離は, \mathcal{M} 上の測地線ではなく, 特徴空間 \mathcal{F} 上で定義される. しかしこれはカーネルトリックを適用することにより, データの空間 \mathcal{X} 上の二変数関数として表すことができる.

以上の考察より, ほとんどの議論は \mathcal{X} と同じ有限次元の空間上で行うことができる. 以下では一般的なカーネル関数について, 本モデルを幾何学的に考察する.

4.7.1 テンソルの表記法

この節では, 双線形写像を表現する行列の一般化であるテンソル [107] を扱う. そこで, 以下のように他の節とは異なる表記法を用いる. あるベクトル空間 V について,

1. V の基底を表すときには下付きの添字を用い, 成分 (基底にかかる係数) には上付きの添字を用いる.
2. このとき, V^* の双対空間として定義されるベクトル空間には 1 と逆の添字を用いる.
3. アインシュタインの規約に従って, 上下の同じ添字による総和については \sum を省略する.
4. 添字には i, j, k, l, m を用い, 上付き文字が数字の場合は冪乗を表す.

例えば, ベクトル空間 V とその双対空間 V^* の元をそれぞれ x, y とすると, 基底 $\{e_i\}_i$ と双対基底 e^j_j ($e^j(e_i) = \delta_i^j$, δ はクロネッカーのデルタ) による成分表示を $x = x^i e_i, y = y_j e^j$ と書き, $y(x) \in \mathbb{R}$ を $(y_j e^j)(x^i e_i) = x^i y_j e^j(e_i) = x^i y_j \delta_i^j = x^i y_i$ と書く. この場合, i, j が上下の添字として出現する式でそれぞれ $\sum_{i=1}^{\dim(V)}$ と $\sum_{j=1}^{\dim(V)}$ が省略されている.

4.7.2 入力画像と基底画像の空間

前述のように \mathcal{F} は高次元空間である. しかし, ϕ の像の空間 $\phi(\mathbb{R}_+^m) = \{\phi(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}_+^m\}$ に限ると, これは m 次元の空間であるといえる. 実際, $\phi(\mathbb{R}_+^m)$ の元は x_1, \dots, x_m の m 個の非負実数で記述することができる. ただし, $\phi(\mathbf{x}) + \phi(\mathbf{y})$ や $a\phi(\mathbf{x}) (a \in \mathbb{R}_+)$ が $\phi(\mathbb{R}_+^m)$ に含まれる, つまり別の $\mathbf{z} \in \mathbb{R}_+^m$ で $\phi(\mathbf{z})$ と表されるとは限らないので, 通常のベクトル空間のような扱いはできない. ここでは $\phi(\mathbb{R}_+)$ に相当する空間を m 次元リーマン多様体 [108, 109] として考えることで, カーネル関数が与える画像間の位相構造の変化を幾何学的に考察する.

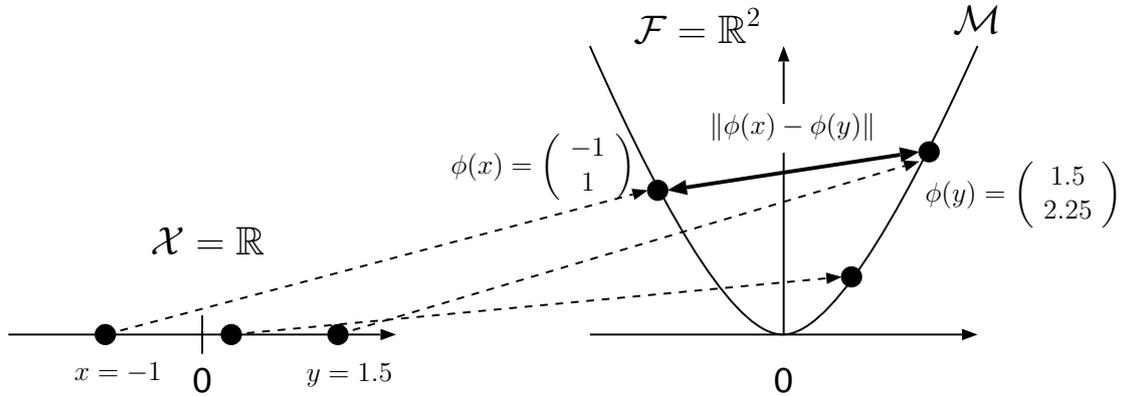


図 4.7: 二次元の特徴空間の例.

$k: \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$ を関数値が常に非負をとる正定値カーネルとする. k を再生核とする再生核ヒルベルト空間 $\mathcal{H}_k = \{f: \mathbb{R}_+^m \rightarrow \mathbb{R}\}$ にノルムから導かれる通常の位相を入れる. 適当な関数系 $\{v_i\}_{i=1}^\infty$ を基底として $f \in \mathcal{H}_k$ を $f = f^i v_i$ と表せるとすると, $\phi: f \mapsto \{f^i\}_{i=1}^\infty$ を局所座標として, \mathcal{H}_k は無限次元の多様体とみなせる. 入力画像と基底画像が占める特徴空間を

$$\mathcal{M} = \{k(\mathbf{x}, \cdot) \mid \mathbf{x} \in \mathbb{R}_+^m\} \quad (4.58)$$

として \mathcal{H}_k と同様の位相を入れる. 再生核ヒルベルト空間の性質より, \mathcal{M} は \mathcal{H}_k に含まれる [110]. LN 型のカーネル関数の場合, $f(t)$ が単調増加ならば, 任意の \mathbf{x} と関数 $k(\mathbf{x}, \cdot)$ が一対一で対応する. この対応 $k(\mathbf{x}, \cdot) \mapsto \mathbf{x}$ を局所座標とすると \mathcal{M} は m 次元多様体となる.

滑らかな曲面の各点において接平面を考えられるように, 可微分多様体の各点 p に接空間 $T_p \mathcal{M}$ を対応させることができる. \mathcal{M} の接ベクトルを定義するために, \mathcal{M} 上の関数について考える. \mathcal{M} の開集合 N 上で微分可能な実数値関数の集合を $\mathfrak{F}(\mathcal{M})$ で表す. 同様に $p \in \mathcal{M}$ の近傍で微分可能な関数の集合を $\mathfrak{F}(p)$ で表す. p における接ベクトルは $h \in \mathfrak{F}(p)$ の方向微分によって定義される. すなわち, $c(0) = p$ となる \mathcal{M} 上の曲線 $c: \mathbb{R} \rightarrow \mathcal{M}$ を用いて接ベクトル $T \in T_p \mathcal{M}$ を

$$Th = \left. \frac{d}{dt} \right|_{t=0} h(c(t)) \quad (4.59)$$

とする. この場合の「方向」は, \mathbf{x} が微小に変化したときの関数 $k(\mathbf{x}, \cdot)$ の変化を表している. それ以外の変化では \mathcal{M} を逸脱するためである. 従って接ベクトルは $T = t^i \left(\frac{\partial}{\partial x_i} \right)_p$ のように m 次元ベクトルとして成分表示できる. ここで基底ベクトル

ル $\left(\frac{\partial}{\partial x_i}\right)_p$ を

$$\left(\frac{\partial}{\partial x_i}\right)_p h = \frac{\partial}{\partial x_i} h(p) \quad (4.60)$$

と定義する。「方向微分」と「方向」を同一視するならば、 $p = k(\mathbf{x}_0, \cdot)$ 上の接ベクトル $\left(\frac{\partial}{\partial x_i}\right)_p$ に関数 $\frac{\partial}{\partial x_i} k(\mathbf{x}_0, \cdot)$ が対応する。

M 上の全ての接ベクトルの集合 $TM = \bigcup_p T_p M$ を M の接束と呼ぶ。また、 M 上の点からその点での接ベクトルへの対応 $M \rightarrow TM$ をベクトル場と呼び、 M 上のベクトル場の集合を $\mathfrak{X}(M)$ とする。例えば $U(p) = \left(\frac{\partial}{\partial x_i}\right)_p$ はベクトル場である。この $U(p)$ を $\frac{\partial}{\partial x_i}$ または ∂_i で表す。一般のベクトル場は $H(p) = h^i(p)\partial_i$ ($h^i \in \mathfrak{F}(M)$) のように関数による重み付き和で表すことができるので、 $\mathfrak{X}(M)$ は $\mathfrak{F}(M)$ 上の加群である。

接ベクトルと同様に、ベクトル場 $\frac{\partial}{\partial x_i}$ と、 \mathbf{x} を固定しない関数 $\frac{\partial}{\partial x_i} k(\mathbf{x}, \cdot)$ を同一視することができる。この同一視により、次のようにベクトル場の基底に内積を与える。

$$\left\langle \frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j} \right\rangle_{\mathfrak{X}(M)} = \left\langle \frac{\partial}{\partial x_i} k(\mathbf{x}, \cdot), \frac{\partial}{\partial x_j} k(\mathbf{x}, \cdot) \right\rangle_{\mathcal{H}_k} = k_{(i)(j)}(\mathbf{x}, \mathbf{x}) \quad (4.61)$$

$\frac{\partial}{\partial x_i} k(\mathbf{x}, \cdot), \frac{\partial}{\partial x_j} k(\mathbf{x}, \cdot)$ は再生核ヒルベルト空間 \mathcal{H}_k に属しており、その内積を用いて定義していることから、 $\langle \cdot, \cdot \rangle_{\mathfrak{X}(M)}$ は内積の公理（双線形性，対称性と正定値性）を満たす。

以上より、計量テンソル場 $g_{ij} = \langle \partial_i, \partial_j \rangle_{\mathfrak{X}(M)}$ が得られ、これによって M にリーマン計量が入る。計量が入ることによって、カーネル関数 k から導かれる特徴空間の局所的な構造を考えられる。画像の微小な変化を $d\mathbf{x} = dx^i \partial_i$ とおくと、画像の特徴の変化は M 上の接ベクトルの二乗ノルム $ds^2 = dx^i dx^j g_{ij}$ で与えられる。

ここでは再生核ヒルベルト空間 \mathcal{H}_k の部分集合として定義しているが、冒頭で述べた ϕ の像空間 $\phi(\mathbb{R}_+^m)$ 上の線素を考えても同様の計量が定義できる。 \mathbf{x} を \mathbb{R}^m の標準基底 $\{e_i\}_i$ を用いて $\mathbf{x} = x^i e_i$ と成分表示する。同様に、特徴ベクトル $\phi = \phi(\mathbf{x})$ を、 $\langle f_i, f_j \rangle_{\mathcal{F}} = \delta_{ij}$ となる \mathcal{F} の基底 $\{f_i\}_i$ を用いて $\phi = \phi^i f_i$ と成分表示できるとすると、各成分 ϕ^i はそれぞれ \mathbf{x} の関数である。このとき ϕ^i の全微分は

$$d\phi^i = \frac{\partial \phi^i}{\partial x_j} dx^j \quad (4.62)$$

となる。

$$ds^2 = d\phi^k d\phi_k = \frac{\partial \phi^k}{\partial x_i} dx^i \frac{\partial \phi_k}{\partial x_j} dx^j \quad (4.63)$$

$d\phi^k d\phi_k$ を $d\phi = d\phi^i f_i$ の二乗ノルム $\langle d\phi, d\phi \rangle_{\mathcal{F}}$ と考えると,

$$\begin{aligned} ds^2 &= \langle d\phi^k f_k, d\phi^l f_l \rangle_{\mathcal{F}} = \left\langle \frac{\partial \phi^k}{\partial x_i} dx^i f_k, \frac{\partial \phi^l}{\partial x_j} dx^j f_l \right\rangle_{\mathcal{F}} \\ &= dx^i dx^j \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} \langle \phi^k f_k, \phi^l f_l \rangle_{\mathcal{F}} = dx^i dx^j \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} \langle \phi, \phi \rangle_{\mathcal{F}} \\ &= dx^i dx^j k_{(i)(j)}(\mathbf{x}, \mathbf{x}) \end{aligned} \quad (4.64)$$

となるので, $k_{(i)(j)}(\mathbf{x}, \mathbf{x})$ は確かに $\phi(\mathbb{R}_+^m) \subset \mathcal{F}$ の計量を与える. 式(4.64)において, 第三の等号では内積と微分の線型性を用いた.

特徴空間のより大域的な構造を考えるため, M の曲率を求める. 曲率は計量テンソル場の微分から求められるから, カーネル関数のより高い次数の偏微分が含まれる. 特に, 式(4.32)を用いると, クリストッフエル記号は

$$\Gamma_{ij}^k = g^{kl} K_{(ij)(l)}(\mathbf{x}, \mathbf{x}) \quad (4.65)$$

のように求められる. ここで g^{ij} は g_{ij} の逆行列で, l についてアインシュタインの規約が適用されている. 具体的な曲率テンソル場については後述する.

4.7.3 予測画像の空間

特徴空間上の予測画像は $\phi(\mathbf{a}_i) (i = 1, \dots, n)$ の線形和で与えられる. 前節で述べた通り $\phi(\mathbb{R}_+^m)$ は加群の構造を持たないので, 新たに予測画像の空間を定義する必要がある.

基底画像のセット $A = (\mathbf{a}_1, \dots, \mathbf{a}_n)$ に対する予測画像の特徴空間を

$$\mathcal{N} = \left\{ \sum_{i=1}^n y_i k(\mathbf{a}_i, \cdot) \mid y_i \in \mathbb{R}_+ \right\} \quad (4.66)$$

で定義する. A を固定すると, \mathbb{R}_+ (半群) 上の加群となり, y_1, \dots, y_n が局所座標を与えるので n 次元多様体とみなせる.

\mathcal{N} の計量は, M の場合と同様に f の方向微分

$$Xf = \frac{d}{dt} \Big|_{t=0} f \left(\sum_i k(\mathbf{a}_i, \cdot) y_i(t) \right) \quad (4.67)$$

を考え, $\frac{d}{dy_i}$ を $k(\mathbf{a}_i, \cdot)$ と同一視することで

$$g_{ij} = \langle k(\mathbf{a}_i, \cdot), k(\mathbf{a}_j, \cdot) \rangle_{\mathcal{H}_k} = k(\mathbf{a}_i, \mathbf{a}_j) \quad (4.68)$$

とする. \mathcal{N} の計量は座標に対して定数となるので, \mathcal{N} は平坦である.

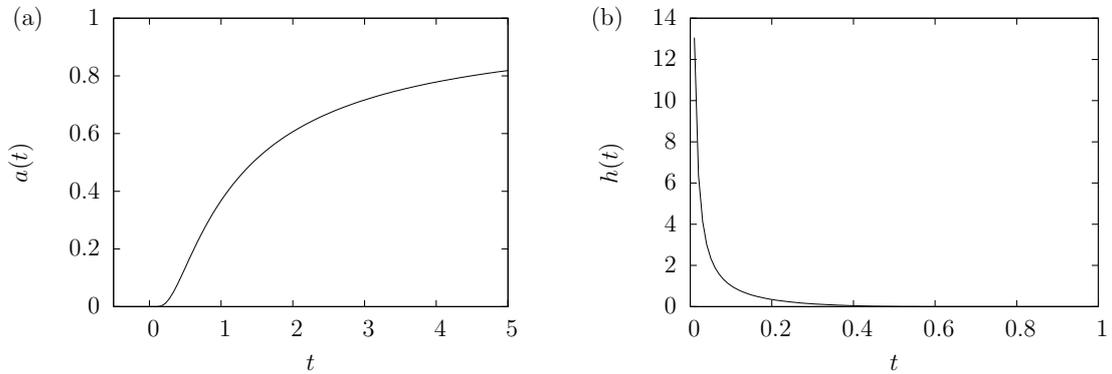


図 4.8: 罰則項. (a) $t \leq 0$ で 0 をとり, C^∞ となる関数 $a(t)$. (b) 罰則項 $h(t)$.

4.7.4 応答特性

ここで \mathcal{M} から \mathcal{N} への写像を考えることで, 入力画像 \mathbf{x} に対するニューロン応答 \mathbf{y} の特性を考える. カーネル関数が正定値で \mathbf{y} の非負制約がなければ, \mathbf{x} に対する \mathbf{y} の値を求める問題は, 凸二次計画として解析的に解ける. すなわち, 目的関数 E を \mathbf{y} の関数としてみたときの極値を取ればよい. 一方, $\mathbf{y} \in \mathbb{R}_+^n$ の制約条件では, E の極値が負の成分を持つときに (\mathbb{R}_+^n を開集合と考えているため), E の最小値が存在しないことになる.

そこで, 任意の $\epsilon > 0$ について $t > \epsilon$ のとき 0 で, $t \rightarrow 0$ のとき正の無限大に発散するような罰則項を考える. まず, 罰則項を $t = \epsilon$ 付近で滑らかに適応するために, 以下の関数を考える.

$$a(t) = \begin{cases} 0 & (t \leq 0) \\ e^{-\frac{1}{t}} & (t > 0) \end{cases}. \quad (4.69)$$

この関数は図 4.8(a) のような形状を持ち, C^∞ 級になることが知られている. また, $t > 0$ について

$$a'(t) = \frac{1}{t^2} e^{-\frac{1}{t}} \quad (4.70)$$

$$a''(t) = \frac{1 - 2t}{t^4} e^{-\frac{1}{t}} \quad (4.71)$$

より, $t \leq \frac{1}{2}$ について $a''(t) \geq 0$ の凸関数である. ところで,

$$(f(t)g(t))'' = f''(t)g(t) + 2f'(t)g'(t) + f(t)g''(t) \quad (4.72)$$

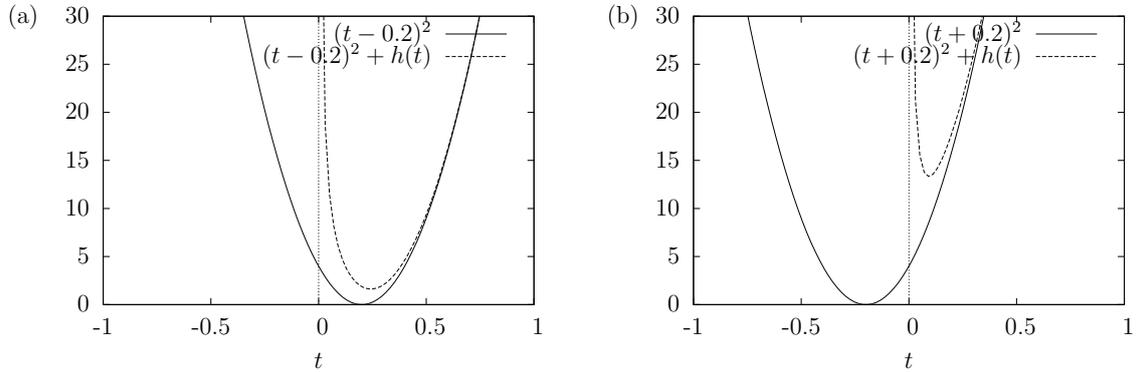


図 4.9: 罰則項による目的関数の変化. (a) と (b) はそれぞれ元の目的関数が $x > 0$ で最小となる場合と $x < 0$ で最小となる場合の例.

より, 2階微分までが非負である関数の積は凸である. 関数 $\frac{1}{t}$ はこのような性質を持っているので, $a(t)$ の $t \leq \frac{1}{2}$ の部分を反転して $\frac{1}{t}$ を重み付けした関数

$$h(t) = a \left(\frac{1}{2} \left(1 - \frac{t}{\epsilon} \right) \right) \frac{1}{t} \quad (4.73)$$

は図 4.8(b) に示すように $t > 0$ について凸で, $t \rightarrow 0$ のとき $h(t) \rightarrow \infty$ である. 従って, 目的関数 E に $\sum_i h(y_i)$ を罰則項として加えることで, 最適化問題の凸性を保ったまま $y_i > 0$ の制約を課することができる. この制約により, E の極値が \mathbb{R}_+^n の外部にあるとき, \mathbf{y} は \mathbb{R}_+^n の境界付近に存在することになる. 図 4.9 は 1 変数の場合の目的関数を示している. この境界付近に対応する \mathcal{N} の部分集合を

$$\mathcal{N}_\epsilon = \left\{ \sum_{i=1}^n k(\mathbf{a}_i, \cdot) y_i \mid \forall i, 0 < y_i < \epsilon \right\} \quad (4.74)$$

とする.

写像 $\Psi: \mathcal{M} \rightarrow \mathcal{N}$ を

$$\Psi(p) = \operatorname{argmin}_q d_{\mathcal{H}_k}^2(p, q) + h(q) \quad (4.75)$$

とする. これは \mathcal{M} 上の点, つまりある画像の特徴ベクトルに対して特徴空間上の予測画像を対応させる写像である. ここで,

$$d_{\mathcal{H}_k}^2(p, q) = \langle p, p \rangle_{\mathcal{H}_k} - 2 \langle p, q \rangle_{\mathcal{H}_k} + \langle q, q \rangle_{\mathcal{H}_k}, \quad (4.76)$$

$$h(q) = \sum_i \frac{c_\epsilon(y_i)}{y_i} \quad (4.77)$$

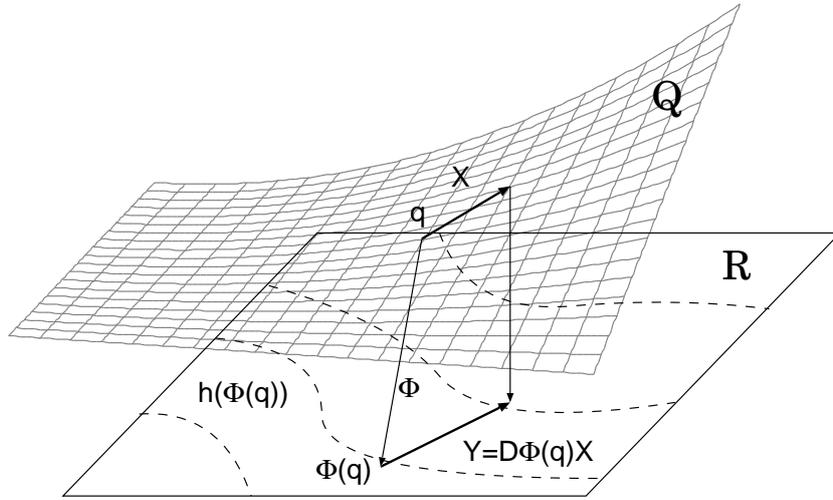


図 4.10: 写像 $\Phi : Q \rightarrow R$ の微分.

である. $\Psi^{-1}(\mathcal{N} \setminus \mathcal{N}_\epsilon)$ においては式 (4.75) は凸二次計画なので, \mathbf{y} を \mathcal{N} の局所座標と考えると

$$\mathbf{y}^i = (\psi_i \circ \Psi \circ \varphi^{-1})(\mathbf{x}) = L^{ij} k(\mathbf{a}_j, \mathbf{x}) \quad (4.78)$$

と表せる. ここで L^{ij} は A のグラム行列 $L_{ij} = \{k(\mathbf{a}_i, \mathbf{a}_j)\}_{ij}$ の逆行列である. また φ は \mathcal{M} の局所座標を与える写像, ψ_i は \mathcal{N} の局所座標の第 i 成分を与える写像である. \mathcal{N}_ϵ の範囲では $\Psi : \mathcal{M} \rightarrow \mathcal{N}$ を解析的に表現することが困難であるため, 以降では $\Psi^{-1}(\mathcal{N} \setminus \mathcal{N}_\epsilon)$ の範囲に限定して議論する.

一般に多様体から多様体への写像 $\Phi : Q \rightarrow R$ の $q \in Q$ における微分 $D\Phi(q)$ は,

$$(D\Phi(q)(X))h = X(h \circ \Phi), \quad X \in T_q Q, \quad h \in \mathfrak{F}(\Phi(q)) \quad (4.79)$$

と定義される [108]. 直感的には図 4.10 のように, R 上の関数 h を, Φ を介して間接的に方向微分する作用 Y を考えることで, Q 上の接ベクトル X を R 上の接ベクトル $Y = D\Phi(q)X$ に写す写像が Φ の微分である. 従って写像 Ψ の微分 $D\Psi$ は

$$\begin{aligned} D\Psi \left(\frac{d}{dx_i} \right) &= \sum_j \frac{d}{dx_i} (\psi_j \circ \Psi \circ \varphi^{-1})(\mathbf{x}) \frac{d}{dy_j} \\ &= L^{jk} k_{(i)(j)}(\mathbf{a}_k, \mathbf{x}) \frac{d}{dy_j} \end{aligned} \quad (4.80)$$

となる。また、この微分をもとに \mathcal{M} の計量として新たに \mathcal{N} からの誘導計量

$$\begin{aligned}
 g_{ij} &= \left\langle \frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j} \right\rangle_{\mathfrak{X}(\mathcal{M})} \\
 &= \left\langle L^{st} k_{() (i)}(\mathbf{a}_s, \mathbf{x}) \frac{\partial}{\partial y_t}, L^{uv} k_{() (j)}(\mathbf{a}_u, \mathbf{x}) \frac{\partial}{\partial y_v} \right\rangle_{\mathfrak{X}(\mathcal{N})} \\
 &= L^{st} k_{() (i)}(\mathbf{a}_s, \mathbf{x}) L^{uv} k_{() (j)}(\mathbf{a}_u, \mathbf{x}) \left\langle \frac{\partial}{\partial y_t}, \frac{\partial}{\partial y_v} \right\rangle_{\mathfrak{X}(\mathcal{N})}
 \end{aligned} \tag{4.81}$$

が得られる。式(4.68)を代入すると、 $L^{ik} L_{jk} = \delta_j^i$ より

$$\begin{aligned}
 g_{ij} &= L^{st} L^{uv} L_{tv} k_{() (i)}(\mathbf{a}_s, \mathbf{x}) k_{() (j)}(\mathbf{a}_u, \mathbf{x}) \\
 &= L^{st} k_{() (i)}(\mathbf{a}_s, \mathbf{x}) k_{() (j)}(\mathbf{a}_t, \mathbf{x})
 \end{aligned} \tag{4.82}$$

となる。

以上で \mathcal{M} に対して二つの計量が定義された。式(4.61)の計量テンソル場を $G_{\mathcal{M}}^1$ 、式(4.81)の計量テンソル場を $G_{\mathcal{M}}^2$ とおく。 $G_{\mathcal{M}}^1$ は、カーネル関数を選択した時点が決まる、入力画像の計量である。これによってカーネル関数自体の性質を知ることができる。一方 $G_{\mathcal{M}}^2$ は、 \mathcal{M} 上の線素を \mathcal{N} に投影して測った計量といえる。つまり $G_{\mathcal{M}}^2$ によって、画像の微小な変化に対して予測画像がどの程度変化するかを、特徴空間上で測ることができる。これは基底 A に依存する。このため、モデルが学習によって獲得した画像の位相（どの画像がどの画像に近い）を $G_{\mathcal{M}}^2$ によって知ることができる。

4.7.5 具体例

ここまでは、カーネル関数の具体的な形状に一切依存しない、一般的なモデルの幾何学的構造について述べてきた。この節では、実際に数値実験に用いたカーネル関数について述べる。

$K(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}^T \mathbf{y})$ の場合、計量テンソル場 $G_{\mathcal{M}}^1$ の成分表示は

$$g_{ij} = f''(s^2) x_i x_j + f'(s^2) \delta_{ij} \tag{4.83}$$

となる。ここで $s^2 = \mathbf{x}^T \mathbf{x}$ とおいた。 δ_{ij} は単位行列なので、 g_{ij} は重複度1の固有値 $f''(s^2) s^2 + f'(s^2)$ と重複度 $n-1$ の固有値 $f'(s^2)$ を持ち、対応する固有空間はそれぞれ \mathbf{x} の張る1次元空間と残りの $n-1$ 次元空間であることがわかる。このことから g_{ij} の逆行列

$$g^{ij} = -\frac{f''(s^2)}{f'(s^2)(f'(s^2) + f''(s^2)s^2)} x^i x^j + \frac{1}{f'(s^2)} \delta^{ij} \tag{4.84}$$

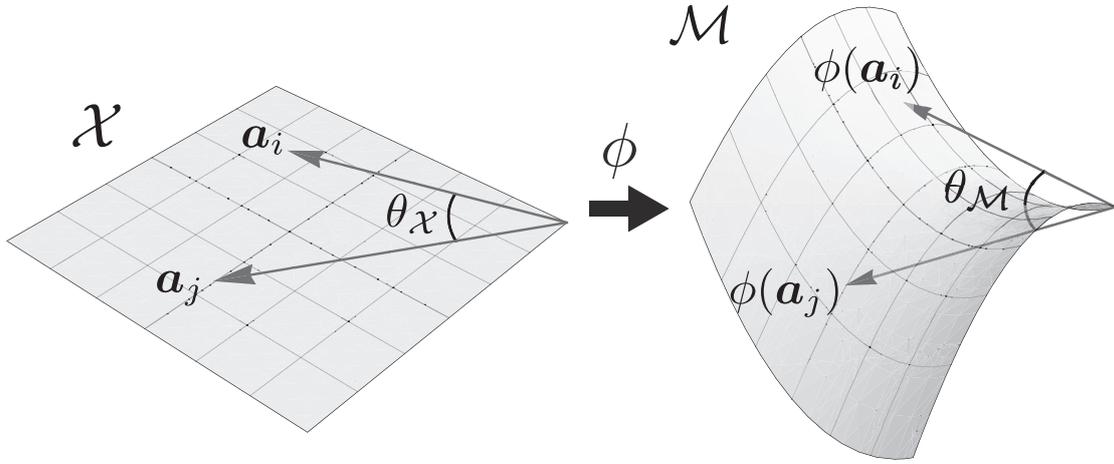


図 4.11: 負曲率を持つ特徴空間への写像. \mathcal{X} は入力と基底の空間, \mathcal{M} は特徴ベクトルの多様体を表す. $\theta_{\mathcal{X}}$ は \mathcal{X} における二つの基底 $\mathbf{a}_1, \mathbf{a}_2$ の成す角, $\theta_{\mathcal{M}}$ はそれらの特徴ベクトルの成す角を表す. 負曲率の場合, 空間の体積要素はユークリッド空間のそれよりも大きくなる.

が得られる. 以上から曲率テンソル場を求めると,

$$\begin{aligned}
 R_{ijk}^l &= \frac{f'(s^2)f^{(3)}(s^2) - f''(s^2)^2}{f'(s^2)^2} (\delta_j^l x_i x_k - \delta_i^l x_j x_k) \\
 &+ \frac{f'(s^2)f^{(3)}(s^2) - f''(s^2)^2}{f'(s^2)(f'(s^2) + f''(s^2)s^2)} (\delta_{ik} x_j x^l - \delta_{jk} x_i x^l) \\
 &+ \frac{f''(s^2)}{f'(s^2)} (\delta_j^l \delta_{ik} - \delta_i^l \delta_{jk})
 \end{aligned} \tag{4.85}$$

となる.

局所座標 \mathbf{x} に対して各成分が一定となるベクトル場 \mathbf{u}, \mathbf{v} を考える. \mathcal{M} 上の各点 p について, \mathbf{u}, \mathbf{v} が張る接空間の 2次元部分空間 $\Sigma_p \subset T_p \mathcal{M}$ が決まる. 写像 $p \mapsto \Sigma_p$ を Σ とする. この Σ の断面曲率 (スカラー場) を求める. Σ が決まれば, 断面曲率は基底の取り方に依存しない.

$$\mathbf{u}' = \mathbf{u} - \frac{\mathbf{u}^T \mathbf{v}}{\mathbf{v}^T \mathbf{v}} \mathbf{v} \tag{4.86}$$

とすることで $\mathbf{u}'^T \mathbf{v} = 0$ となる Σ の基底を取ることができるので, これを新たに \mathbf{u}, \mathbf{v} とすると, $\mathbf{u}^T \mathbf{v} = 0$ とみなせる. 同様に, 基底を定数倍しても Σ が変化することはないので, $\mathbf{u}^T \mathbf{u} = \mathbf{v}^T \mathbf{v} = 1$ とする. ここで $\mathbf{x}^T \mathbf{y}$ は成分ベクトルの標準内積

を表している。以上を考慮して式を整理すると、各点の断面曲率は

$$\kappa_{\Sigma}(\mathbf{x}) = \frac{-f''(s^2) + \frac{f''(s^2)^2}{f'(s^2)} - f^{(3)}(s^2)((\mathbf{x}^T \mathbf{u})^2 + (\mathbf{x}^T \mathbf{v})^2)}{f'(s^2)(f'(s^2) + f''(s^2)((\mathbf{x}^T \mathbf{u})^2 + (\mathbf{x}^T \mathbf{v})^2))} \quad (4.87)$$

と表せる。分母はベクトル場 \mathbf{u}, \mathbf{v} が Σ 上で作る平行四辺形の面積 $\|u \wedge v\|_{\mathbf{x}(\mathcal{M})}^2$ で、カーネル関数が正定値ならば $\|u \wedge v\|_{\mathbf{x}(\mathcal{M})}^2 \geq 0$ である。従って、分子が常に非正であれば、 \mathcal{M} は至る所で負の曲率を持つ。この場合、図4.11に示すように、カーネル関数によって決定される特徴ベクトルの多様体 \mathcal{M} は、画像の値そのものの空間 \mathbb{R}_+ よりも広がっているといえる。

例えば $f(u) = e^{\alpha u}$ の場合、断面曲率は

$$\kappa_{\Sigma}(\mathbf{x}) = -\frac{e^{-\alpha s^2}}{1 + \alpha((\mathbf{x}^T \mathbf{u})^2 + (\mathbf{x}^T \mathbf{v})^2)} \quad (4.88)$$

となる。これは常に負を取ることから、 \mathcal{M} は上で述べた負曲率多様体の性質を持つ。

4.8 まとめ

皮質領野間の双方向結合の役割とその学習則は、入力信号の生成過程のモデルを内部に持ち、入力信号を最もよく表現するようにそれを修正するという、生成モデルの枠組みで説明される。しかし、勾配法によって予測誤差を最小化するようにネットワークのダイナミクスを導出すると、V1でみられるLN型の非線型性が説明できない。

本モデルでは、カーネル法を援用することで、生成モデルの枠組みに非線型性を導入した。また、生理学的妥当性のため、変数を非負に制約した。V1でみられる expansive な非線型を持つLN型のカーネル関数によって、モデルニューロンの応答はスパースに分布した。これは、LNモデルの前半である線型フィルタの応答 u が0に近いときに出力はほぼ0となり、 u がある程度大きくなると一転して出力が大きくなるためである。また自然画像の学習後、本モデルはV1でみられるガボール型の受容野を獲得した。

スパースコーディングは、V1が網膜像を線型基底のスパースな組み合わせによって表現することを仮定したモデルである。一方、本モデルでは、スパース性を仮定する代わりに非負性と非線型性を仮定しており、スパース性は結果として得られている。非負性の仮定については、活動電位によって情報を伝達するというニューロンの生物学的基盤において、ある程度の妥当性があるといえる。一方、非線型性については、なぜ expansive なのかという問題がある。

この問題を解決するため、本モデルが表現する特徴空間について幾何学的な考察を行った。幾何学的な分析を行う大きな利点は、カーネル関数の微分から特徴空間の計量が直接得られることである。正定値性を確認できないカーネル関数についても、接空間の正定値性を調べることによって、特徴空間の局所的な構造がわかる。

本章の後半では、カーネル関数によって定まる特徴空間の断面曲率を導出した。カーネル関数の expansive な非線型性によって、非負ベクトル空間は、任意の二つのベクトルが殆ど至る所で直交しないという意味において、狭い空間であるといえる。Expansive な非線型性は、この狭い空間において直交基底を抽出するための非ユークリッド的な計量を与えているといえる。

カーネル法が用いられる多くの多変量分析は、関数

$$f_{\mathbf{w}}(\mathbf{x}) = \langle \phi(\mathbf{x}), \mathbf{w} \rangle_{\mathcal{H}} \quad (4.89)$$

を、問題に依存する目的関数に従って最適化しているという点で共通している。特徴空間 \mathcal{H} は一般に無限次元なので、このままでは $\mathbf{w} \in \mathcal{H}$ を表現することができない。そこで \mathbf{w} のノルムに制約をかけることで、 \mathbf{w} を特徴ベクトル $\phi(\mathbf{x}_t)$ ($t = 1, \dots, T$) の線型和として表現している (リプレゼンター定理)。これによって、 $f_{\mathbf{w}}(\mathbf{x})$ はカーネル関数 $k(\mathbf{x}, \mathbf{x}_t)$ の線型和となり、元の問題は T 次元空間における最適化問題となる。例えば、本モデルと同じくカーネル法を用いた教師なし学習であるカーネル主成分分析 (カーネル PCA) は、 $T \times T$ の大きさを持つ (中心化された) グラム行列の固有値問題に帰着される。これは $f_{\mathcal{H}}(\mathbf{x})$ が T 個の全てのサンプル \mathbf{x}_t によって表現されることを意味する。視覚系の入力 は網膜像として常に外界から与えられているため、このような表現は不可能である。カーネル PCA と本モデルの大きな違いは、入力が一定数の基底ベクトル \mathbf{a}_j ($j = 1, \dots, n$) によって表現されることと、それら基底ベクトルが学習によって変化することである。

非線型性によってスパースな表現を獲得するモデルは、Bell ら [111] のものが有名である。これは情報量最大化原理を実現する一つのアルゴリズムである。いま、入力 $\mathbf{x} \in \mathbb{R}^m$ と出力 $\mathbf{y} \in \mathbb{R}^n$ の関係を

$$\mathbf{y} = f(W^T \mathbf{x}) \quad (4.90)$$

のような LN 型の関数で表す。ここで $m = n$ とすると、式 (4.90) は確率変数の変数変換であり、 $p(\mathbf{y}) = |\partial \mathbf{y} / \partial \mathbf{x}|^{-1} p(\mathbf{x})$ となる。このとき出力のエントロピーは

$$H(\mathbf{y}) = H(\mathbf{x}) + E \left[\log \left| \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \right| \right] \quad (4.91)$$

で与えられ、その最大化は第二項の最大化に帰着される。Bell らは $f(u)$ がシグモイド関数のような典型的な非線型性の場合、 $H(\mathbf{y})$ の勾配が \mathbf{y} や $\mathbf{u} = W^T \mathbf{x}$ の簡単な関

数で表せることを利用して anti-Hebb 型の学習則を導いた。例えば $f(u) = \tanh(u)$ の場合、

$$\Delta W \propto \frac{\partial H(\mathbf{y})}{\partial W} = W^{-T} - 2\mathbf{x}\mathbf{y}^T \quad (4.92)$$

となり、原信号が super-Gaussian に従う場合は分離行列を推定できる。従ってこのモデルは、求心性の結合と非線型性（式(4.90)）で出力を計算し、学習は入力と出力の anti-Hebb 則で実現される。このため、Bell らのモデルは遠心性結合の役割までを説明するモデルではない。本モデルは、皮質領野間の双方向結合の計算論的役割を説明することを可能とする生成モデルの枠組みに、非線型性を自然な形で導入したモデルであるといえる。

第5章 結論

脳機能の計測技術の進歩により、現在では脳の様々な特性が明らかになっている。また得られた特性が計算論的にどのような役割を持っているのかも徐々に明らかになりつつある。しかし、非線型に応答するニューロンの特性を適切に記述し、計測することは今もなお困難な課題である。本研究は大きく分けて、以下の二つの課題に取り組んだ。

第二章では、逆相関法に区分線形近似を導入し、非線型なニューロン応答の局所的な特性を分析する手法を提案した。この手法によって、入力空間における局所領域を対象に測定し、応答特性を線型近似によって記述することができることが示された。

第三章と第四章では、ニューロン応答の非線型性を説明するための学習モデルを構築した。複雑型細胞は、様々な空間位相に応答する単純型細胞から興奮性の入力を受けることによって、位相不変な応答を実現していると考えられる。この中には、互いにほぼ逆位相の関係にある受容野の組み合わせも含まれている。このような単純型細胞の応答は明らかに負の相関を持つため、単純型細胞-複雑型細胞間の結合荷重の学習はSCのような線型な生成モデルによる学習則では説明できない。第三章で提案した二階層の階層ベイズモデルでは、第二層において、線型モデルでは獲得できない、複数の Gabor 型基底の組み合わせを獲得できることを示した。ただし、それは複雑型応答を構成するような、空間位相の組み合わせとはならなかった。第四章で提案したモデルでは、カーネル法を援用することで、生成モデルの枠組みに非線型性を導入し、生理学的妥当性のため、変数を非負に制約した。数値実験の結果より、V1 でみられる expansive な非線型を持つ LN 型の非線型性によって、ニューロン応答がスパースになることが確かめられた。またこのとき、学習によって獲得される基底は Gabor 型となった。ヒルベルト空間上の線型作用素から導かれる再生核の理論は、現在ではカーネル法として工学的に重要な役割を担っている。これは短時間で計算できるカーネル関数が、暗に高次元空間での内積を表現することにより、従来の線型な手法の性能を飛躍的に向上させられるためである。本研究ではカーネル法を多変量データの計算のためではなく、ニューロン応答特性の計算論的解釈のために用いた。その結果より、ニューロン応答の非線型性とスパースで冗長性の少ない情報の符号化との間には、再生

核の理論が関係していることが示唆される。

大脳皮質における視覚情報処理は後頭葉の V1 に始まり、腹側経路では物体の形状や色などの処理が、背側経路では主に物体の位置や動きに関する処理が行われる。各経路は解剖学的、機能的に異なるいくつかの領野に分けられ、これらが双方向的に連絡し合って徐々に高次の処理を行うという、階層構造を持っている。各領野のニューロンがどのような特徴に応答するかは、IT に至るまで部分的にはわかってきている。例えば IT において顔に反応するニューロンが報告されている [29]。一方、このような特徴選択性がどのように実現されているかについての理解は限定的である。例えば V1 の単純型細胞の方位選択性は on-center, off-center 型受容野の組み合わせによる Gabor 型受容野によって説明できるが、V4 における曲率の表現 [25] や IT における顔に対する選択性の実現に関するはつきりとした理解は得られていない。電気生理実験によるニューロン応答特性の分析は、上記の問題に対する重要なアプローチとなる。従来の線型逆相関法は、単純型細胞の詳細な受容野形状を機械的に測定できる点で優れた手法であるが、一般の非線型な応答特性には対応できない。第二章で提案した分析手法は、応答特性の（画像平面でなく画像そのものの空間における）局所的な性質を探る新しいアプローチであり、応答特性の勾配を近似的に推定できることから様々な応用が期待できる。例えば、ニューロンの既知の適刺激を開始点として他の適刺激を探索する手法などを考えることができる。

上記のように視覚皮質の構造が固定的である一方、脳は柔軟な可塑性を持っている。片眼遮蔽下で育った仔ネコの例 [3] からわかるように、皮質領野の計算機構は、少なからず後天的な学習によって獲得されているといえる。しかし、特徴選択性を実現する計算機構がどのような学習によって獲得されるのかという問いに対する答えは、計算機構そのものの理解よりもさらに限定的である。本研究で扱った生成モデルは、外界からの入力に基づいて最も効率のよい表現を獲得するという説明を以てこの問いに答え得る大きな枠組みである。このため初期視覚皮質の多くの学習モデルがこの枠組みに基づいている。第四章で提案したモデルでは、非線型性を自然な形で導入するためにカーネル法を援用した。大脳皮質の各領野は解剖学的な構造の違いによって分けられるが、大脳皮質としての基本的な構造は同様である。従って、階層構造の単位となる単一領野の入出力特性を十分に記述できれば、高次領野の複雑なパターン選択性についても同様のモデルの積み重ねで実現できると考えられる。このため最も初期の領野において非線型性も含めた学習理論を構築することに大きな意義があるといえる。

本研究の全体を通して、非線型関数をどのように表現するかが共通の課題となっている。一般的なカーネル法では、リプレゼンター定理を利用し、一般に無限次元の特徴空間のうちデータ数次元の部分空間で十分な表現が得られる。この場合、

非線型関数はデータによって定まる基底関数の線型和で表され、その重みを求めることが目的となる。これはデータ数が有限である工学的な問題には広く応用することが可能である。しかし、視覚系は時々刻々と入力される視覚刺激をオンラインで処理しなければならない。第四章で提案したモデルでは、シナプスの長期的な変化は基底関数そのものの形状を変化させる。このため予め十分に多く用意した基底関数の重みのみを学習するのではなく、少ない個数の基底関数の重みを求めながら、長期的にはデータ全体をよく表現する基底関数の形状を学習することができる。本モデルが脳皮質の学習理論をよく表しているのであれば、高次ニューロン応答やその可塑性を研究することで、多変量で非線型な関数を効率的に表現する方法が得られ、様々な分野に幅広く応用が可能になると考えられる。

謝辞

本研究を進めるにあたり，研究活動に必要な機材を始めとした多大なる御支援を賜り，有益な御助言をいただきました，室蘭工業大学大学院工学研究科 渡部修准教授に心から深く感謝致します。

同 畑中雅彦教授は博士後期課程へ進むきっかけを作ってください，日頃から御助言や御激励を賜りました。厚くお礼申し上げます。

本論文の執筆にあたり，同 佐賀聡人教授には貴重な御助言を賜りました。謹んで謝意を表します。

また，同 鈴木幸司教授，塩谷浩之教授，本田泰准教授，岡田吉史准教授，倉重健太郎助教，寺本渉助教には研究のみならず，様々な面でお世話になりました。心よりお礼申し上げます。

最後に，長い学生生活において様々な面で大きな支えとなり，本研究にとっても少なからずよい刺激を与えてくださった，研究室の先輩や同期，後輩の方々，学会で知り合った他大学の方々，友人や家族，親戚の皆様に心からお礼申し上げます。

参考文献

- [1] P. Dayan and L. F. Abbott. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. The MIT Press, 2005.
- [2] A. Anzai, I. Ohzawa, and R. D. Freeman. Neural mechanisms for processing binocular information. I. Simple cells. *Journal of Neurophysiology*, 82(2):891–908, 1999.
- [3] D. H. Hubel and T. N. Wiesel. The period of susceptibility to the physiological effects of unilateral eye closure in kittens. *Journal of Physiology*, 206(2):419–436, 1970.
- [4] T. N. Wiesel. Postnatal development of the visual cortex and the influence of environment. *Nature*, 299(5884):583–591, 1982.
- [5] D. Hebb. *The Organization of Behavior: A Neuropsychological Theory*. Psychology Press, 1970.
- [6] T. Bliss and T. Lomo. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *The Journal of Physiology*, 232(2):331–356, 1973.
- [7] A. Artola and W. Singer. Long-term potentiation and NMDA receptors in rat visual cortex. *Nature*, 330(6149):649–652, 1987.
- [8] A. Iriki, C. Pavlides, A. Keller, and H. Asanuma. Long-term potentiation in the motor cortex. *Science*, 245(4924):1385–1387, 1989.
- [9] J. C. Hirsch and F. Crepel. Postsynaptic calcium is necessary for the induction of LTP and LTD of monosynaptic EPSPs in prefrontal neurons: An in vitro study in the rat. *Synapse*, 10(2):173–175, 1992.

- [10] S. Dudek and M. Bear. Homosynaptic long-term depression in area CA1 of hippocampus and effects of N-methyl-D-aspartate receptor blockade. *Proceedings of the National Academy of Sciences of the United States of America*, 89(10):4363–4367, 1992.
- [11] E. Bienenstock, L. Cooper, and P. Munro. Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *The Journal of Neuroscience*, 2(1):32–48, 1982.
- [12] L. I. Zhang, H. W. Tao, C. E. Holt, W. A. Harris, and M. Poo. A critical window for cooperation and competition among developing retinotectal synapses. *Nature*, 395(6697):37–44, 1998.
- [13] W. Levy and O. Steward. Temporal contiguity requirements for long-term associative potentiation/depression in the hippocampus. *Neuroscience*, 8(4):791 – 797, 1983.
- [14] H. Markram, J. Lubke, M. Frotscher, and B. Sakmann. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, 275(5297):213–215, 1997.
- [15] G. Bi and M. Poo. Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience*, 18(24):10464–10472, 1998.
- [16] N. Caporale and Y. Dan. Spike timing-dependent plasticity: A hebbian learning rule. *Annual Review of Neuroscience*, 31:25–46, 2008.
- [17] G. E. Schneider. Two visual systems. *Science*, 163(3870):895–902, 1969.
- [18] D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *The Journal of Physiology*, 160:106–154, 1962.
- [19] D. H. Hubel and T. N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, 195(1):215–243, 1968.
- [20] G. C. DeAngelis, I. Ohzawa, and R. D. Freeman. Spatiotemporal organization of simple-cell receptive fields in the cat’s striate cortex. II. Linearity of temporal and spatial summation. *Journal of Neurophysiology*, 69(4):1118–1135, 1993.

- [21] E. H. Adelson and J. R. Bergen. Spatio-temporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2(2):1861, 1985.
- [22] M. Ito and H. Komatsu. Representation of angles embedded within contour stimuli in area V2 of macaque monkeys. *The Journal of Neuroscience*, 24(13):3313–3324, 2004.
- [23] A. Anzai, X. Peng, and D. C. Van Essen. Neurons in monkey visual area V2 encode combinations of orientations. *Nature Neuroscience*, 10(10):1313–1321, 2007.
- [24] J. Hegd e and D. C. Van Essen. Selectivity for complex shapes in primate visual area V2. *The Journal of Neuroscience*, 20(5):1–6, 2000.
- [25] A. Pasupathy and C. E. Connor. Shape representation in area V4: Position-specific tuning for boundary conformation. *Journal of Neurophysiology*, 86(5):2505–2519, 2001.
- [26] A. Pasupathy and C. E. Connor. Population coding of shape in area V4. *Nature Neuroscience*, 5(12):1332–1338, 2002.
- [27] E. Kobatake and K. Tanaka. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology*, 71(3):856–867, 1994.
- [28] E. Kobatake, G. Wang, and K. Tanaka. Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *Journal of Neurophysiology*, 80(1):324–330, 1998.
- [29] D. Perrett, E. Rolls, and W. Caan. Visual neurones responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, 47:329–342, 1982.
- [30] Y. Sugase, S. Yamane, S. Ueno, and K. Kawano. Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, 400(6747):869–873, 1999.
- [31] D. Marr. ビジヨン: 視覚の計算理論と脳内表現. 産業図書, 1987.
- [32] 川人 光男. 脳の計算理論. 産業図書, 1996.

- [33] A. Hodgkin and A. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117(4):500–544, 1952.
- [34] N. Schweighofer, K. Doya, and M. Kawato. Electrophysiological properties of inferior olive neurons: A compartmental model. *Journal of Neurophysiology*, 82(2):804–817, 1999.
- [35] W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5:115–133, 1943.
- [36] 銅谷 賢治. 計算神経科学への招待: 脳の学習機構の理解を目指して. SGC ライブラリ. サイエンス社, 2007.
- [37] K. Doya. What are the computations of the cerebellum, the basal ganglia, and the cerebral cortex? *Neural Networks*, 12:961–974, 1999.
- [38] 銅谷 賢治, 藤井 宏, 伊藤 浩之, 塚田 稔. 脳の情報表現—ニューロン・ネットワーク・数理モデル. 朝倉書店, 2002.
- [39] R. Linsker. From basic network principles to neural architecture: Emergence of spatial-opponent cells. *Proceedings of the National Academy of Sciences*, 83(19):7508–7512, 1986.
- [40] R. Linsker. From basic network principles to neural architecture: Emergence of orientation-selective cells. *Proceedings of the National Academy of Sciences*, 83(21):8390–8394, 1986.
- [41] R. Linsker. From basic network principles to neural architecture: Emergence of orientation columns. *Proceedings of the National Academy of Sciences*, 83(22):8779–8783, 1986.
- [42] H. Barlow. Unsupervised learning. *Neural Computation*, 1:295–311, 1989.
- [43] S. Becker and S. Becker. Unsupervised learning with global objective functions, 2003.
- [44] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, 1994.
- [45] B. A. Olshausen and D. J. Field. Natural image statistics and efficient coding. *Network: Computation in Neural Systems*, 7(2):333–339, 1996.

- [46] B. A. Olshausen. Learning linear, sparse, factorial codes. *A.I. Memo 1580, MIT Artificial Intelligence Lab*, 1996.
- [47] B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, 37(23):3311–3325, 1997.
- [48] B. Olshausen. *Principles of Image Representation in Visual Cortex*, In L. M. Chalupa and J. S. Werner, editors, *The Visual Neurosciences*, chapter 108. 2003.
- [49] 銅谷 賢治, 阪口 豊, 五味 裕章, 川人 光男. 脳の計算機構—ボトムアップ・トップダウンのダイナミクス. 朝倉書店, 2005.
- [50] 川人 光男, 乾 敏郎. 視覚大脳皮質の計算理論. 電子情報通信学会論文誌 *D*, 73(8):p1111–1121, 1990.
- [51] R. P. N. Rao and D. H. Ballard. Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87, 1999.
- [52] R. P. N. Rao. An optimal estimation approach to visual perception and learning. *Vision Research*, 39(11):1963–1989, 1999.
- [53] H. Yokoyama and O. Watanabe. A multi-layer predictive coding model: A comparison between the model property and the inferotemporal cortex. *Neuroscience Research*, 68(1):e322–e323, 2010.
- [54] 横山 裕樹, 渡部 修. 多階層予測符号化モデルに基づくパターンの表現. 電子情報通信学会論文誌 *D*, J94-D(1):384–394, 2011.
- [55] A. Hyvärinen, P. Hoyer, and M. Inki. Topographic ICA as a model of V1 receptive fields. In *Proceedings of the International Joint Conference on Neural Networks*, volume 4, pages 83–88, 2000.
- [56] A. Hyvärinen and P. O. Hoyer. A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images. *Vision Research*, 41(18):2413–2423, 2001.
- [57] A. Hyvärinen, P. O. Hoyer, and M. Inki. Topographic independent component analysis. *Neural Computation*, 13(7):1527–1558, 2001.

- [58] A. Hyvärinen and P. O. Hoyer. Topographic independent component analysis as a model of V1 organization and receptive fields. *Neurocomputing*, 38-40:1307–1315, 2001.
- [59] Y. Karklin and M. S. Lewicki. Learning higher-order structures in natural images. *Network: Computation in Neural Systems*, 14(3):483–499, 2003.
- [60] Y. Karklin and M. S. Lewicki. A hierarchical bayesian model for learning nonlinear statistical regularities in nonstationary natural signals. *Neural Computation*, 17(2):397–423, 2005.
- [61] M. Seeck, C. M. Michel, N. Mainwaring, R. Cosgrove, H. Blume, J. Ives, T. Landis, and D. L. Schomer. Evidence for rapid face recognition from human scalp and intracranial electrodes. *NeuroReport*, 8(12), 1997.
- [62] E. D. Adrian and Y. Zotterman. The impulses produced by sensory nerve-endings: Part II. The response of a single end-organ. *The Journal of Physiology*, 61(2):151–171, 1926.
- [63] A. T. Roussin, J. D. Victor, J. Chen, and P. M. Di Lorenzo. Variability in responses and temporal coding of tastants of similar quality in the nucleus of the solitary tract of the rat. *Journal of Neurophysiology*, 99(2):644–655, 2008.
- [64] H. M. Sakai, K. Naka, and M. J. Korenberg. White-noise analysis in visual neuroscience. *Visual Neuroscience*, 1(3):287–296, 1988.
- [65] E. J. Chichilnisky. A simple white noise analysis of neuronal light responses. *Network: Computation in Neural Systems*, 12(2):199–213, 2001.
- [66] E. P. Simoncelli, J. P. Paninski, J. Pillow, and O. Schwartz. *Characterization of Neural Responses with Stochastic Stimuli*, In M. S. Gazzaniga, editor, *The Cognitive Neurosciences III*. Bradford Books, 2004.
- [67] A. Recio-Spinoso and P. v. Dijk. Analysis of responses to noise in the ventral cochlear nucleus using wiener kernels. *Hearing Research*, 216-217(0):7–18, 2006.
- [68] A. J. Ahumada and J. Lovell. Stimulus features in signal detection. *Journal of The Acoustical Society of America*, 49:1751–1756, 1971.

- [69] P. Neri. Estimation of nonlinear psychophysical kernels. *Journal of Vision*, 4(2):82–91, 2004.
- [70] D. L. Ringach, M. J. Hawken, and R. Shapley. Receptive field structure of neurons in monkey primary visual cortex revealed by stimulation with natural image sequences. *Journal of Vision*, 2(1):12–24, 2002.
- [71] J. Rapela, J. M. Mendel, and N. M. Grzywacz. Estimating nonlinear receptive fields from natural images. *Journal of Vision*, 6(4):441–474, 2006.
- [72] H. Yu and V. R. de Sa. Nonlinear reverse correlation with synthesized naturalistic noise. *Neurocomputing*, 58–60(0):909–913, 2004.
- [73] D. Q. Nykamp and D. L. Ringach. Full identification of a linear-nonlinear system via cross-correlation analysis. *Journal of Vision*, 2(1):1–11, 2002.
- [74] A. Aizerman, E. M. Braverman, and L. I. Rozoner. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, 25:821–837, 1964.
- [75] 赤穂 昭太郎. カーネル多変量解析—非線形データ解析の新しい展開. 岩波書店, 2008.
- [76] J. Shawe-Taylor, N. Cristianini, and 大北 剛. カーネル法によるパターン解析. 共立出版, 2010.
- [77] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20:273–297, 1995.
- [78] K. Sasaki and I. Ohzawa. Internal spatial organization of receptive fields of complex cells in the early visual cortex. *Journal of Neurophysiology*, 98(3):1194–1212, 2007.
- [79] 横山 裕樹, 渡部 修. 区分線形性を考慮したニューロン応答特性の分析手法. In 電子情報通信学会技術報告, NC2010-191, 2011.
- [80] P. Julián, M. Jordán, and A. Desages. Canonical piecewise-linear approximation of smooth functions. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 45(5):567–571, 1998.

- [81] C. C. Pack, B. R. Conway, R. T. Born, and M. S. Livingstone. Spatiotemporal structure of nonlinear subunits in macaque visual cortex. *Journal of Neuroscience*, 26(3):893–907, 2006.
- [82] D. J. Heeger. Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9(2):181–197, 1992.
- [83] M. Schetzen. *The Volterra and Wiener Theories of Nonlinear Systems*. Krieger Pub., 2006.
- [84] D. L. Ringach, G. Sapiro, and R. Shapley. A subspace reverse-correlation technique for the study of visual neurons. *Vision Research*, 37(17):2455–2464, 1997.
- [85] H. Yokoyama and O. Watanabe. A hierarchical Bayesian model of the early visual cortex. In *The Proceedings of the 21st Annual Conference of the Japanese Neural Network Society*, pages 60–61, 2011.
- [86] 横山 裕樹, 渡部 修. 階層ベイズモデルによる複雑型細胞の偶対称な応答特性の獲得. In 電子情報通信学会技術報告, NC2011-130, 2012.
- [87] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.
- [88] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *Proceedings of Neural Information Processing Systems*, pages 556–562. MIT Press, 2000.
- [89] P. O. Hoyer. Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, 5:1457–1469, 2004.
- [90] H. Attias. Inferring parameters and structure of latent variable models by variational bayes. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pages 21–30. Morgan Kaufmann Publishers, 1999.
- [91] V. Smidl and A. Quinn. *The Variational Bayes Method in Signal Processing*. Springer Berlin Heidelberg, 2005.
- [92] D. D. Stettler, A. Das, J. Bennett, and C. D. Gilbert. Lateral connectivity and contextual interactions in macaque primary visual cortex. *Neuron*, 36(4):739–750, 2002.

- [93] M. Mancini, B. C. Madden, and R. C. Emerson. White noise analysis of temporal properties in simple receptive fields of cat cortex. *Biological Cybernetics*, 63(3):209–219, 1990.
- [94] C. Bahlmann, B. Haasdonk, and H. Burkhardt. On-line handwriting recognition with support vector machines — a kernel approach. In *Proceedings of the 8th IWFHR*, pages 49–54, 2002.
- [95] D. Decoste and B. Schölkopf. Training invariant support vector machines. *Machine Learning*, 46(1–3):161–190, 2002.
- [96] B. Haasdonk and D. Keysers. Tangent distance kernels for support vector machines. In *Proceedings of the 16th ICPR*, pages 864–868, 2002.
- [97] P. J. Moreno, P. Ho, and N. Vasconcelos. A Kullback-Leibler divergence based kernel for SVM classification in multimedia applications. In S. Thrun, L. K. Saul, and B. Schölkopf, editors, *Proceedings of Neural Information Processing Systems*. MIT Press, 2003.
- [98] H. Shimodaira, K. Noma, M. Nakai, and S. Sagayama. Dynamic time-alignment kernel in support vector machine. In *Proceedings of Neural Information Processing Systems*. MIT Press, 2001.
- [99] B. Haasdonk. Feature space interpretation of SVMs with indefinite kernels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):482–492, 2005.
- [100] W. Noll. Euclidean geometry and Minkowskian chronometry. *The American Mathematical Monthly*, 71(2):129–144, 1964.
- [101] E. Pekalska, P. Paclik, and R. P. W. Duin. A generalized kernel approach to dissimilarity-based classification. *J. Mach. Learn. Res.*, 2:175–211, 2002.
- [102] E. Pekalska, A. Harol, R. P. Duin, B. Spillmann, and H. Bunke. Non-Euclidean or non-metric measures can be informative. In D. Yeung, J. Kwok, A. Fred, F. Roli, and D. Ridder, editors, *Structural, Syntactic, and Statistical Pattern Recognition*, volume 4109 of *Lecture Notes in Computer Science*, pages 871–880. Springer Berlin Heidelberg, 2006.
- [103] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68(3):337–404, 1950.

- [104] B. Chapman and M. Stryker. Development of orientation selectivity in ferret visual cortex and effects of deprivation. *The Journal of Neuroscience*, 13(12):5251–5262, 1993.
- [105] N. J. Priebe and D. Ferster. Mechanisms underlying cross-orientation suppression in cat visual cortex. *Nature Neuroscience*, 9(4):552–561, 2006.
- [106] B. Li, J. K. Thompson, T. Duong, M. R. Peterson, and R. D. Freeman. Origins of cross-orientation suppression in the visual cortex. *Journal of Neurophysiology*, 96(4):1755–1764, 2006.
- [107] A. I. Borisenko, I. E. Tarapov, Mathematics, and R. A. Silverman. *Vector and Tensor Analysis with Applications*. Dover Publications, 1979.
- [108] 酒井 隆. リーマン幾何学. 裳華房, 1992.
- [109] 松本 幸夫. 多様体の基礎. 東京大学出版会, 1988.
- [110] 斎藤 三郎. 再生核の理論入門. 牧野書店, 2002.
- [111] A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.

研究業績

学術論文

1. 横山 裕樹, 渡部 修. 多階層予測符号化モデルに基づくパターンの表現. 電子情報通信学会論文誌 *D*, J94-D(1):384–394. 2011.

学会発表

1. Yokoyama, Hiroki & Watanabe, Osamu. A multi-layer predictive coding model: A comparison between the model property and the inferotemporal cortex. *Neuroscience Research*, 68(1):e322–e323. 2010.
2. 横山 裕樹, 渡部 修. 区分線形性を考慮したニューロン応答特性の分析手法. 電子情報通信学会技術報告, NC2010-191. 2011.
3. Yokoyama, Hiroki & Watanabe, Osamu. A hierarchical Bayesian model of the early visual cortex. *The Proceedings of the 21st Annual Conference of the Japanese Neural Network Society*, 60–61. 2011.
4. 横山 裕樹, 渡部 修. 階層ベイズモデルによる複雑型細胞の偶対称な応答特性の獲得. 電子情報通信学会技術報告, NC2011-130. 2012.
5. Yokoyama, Hiroki & Watanabe, Osamu. Nonnegative predictive coding with kernels. *The Proceedings of the 22nd Annual Conference of the Japanese Neural Network Society*, P1-7. 2012.