

2020 年度修士論文

NMF とバギングを用いたカルシウムイメージングデータのクラスタリング方法の一検討

5319E056-6 永山 瑞生

2021 年 1 月 11 日

早稲田大学 先進理工学研究科
電気・情報生命専攻
情報学習システム研究室

修士論文概要書

Summary of Master's Thesis

Date of submission: 02/29/2020

専攻名 (専門分野) Department	電気・情報生命	氏名 Name	永山 瑞生	指導教員 Advisor	村田 昇
研究指導名 Research guidance	情報学習システム	学籍番号 Student ID number	5319E056-6		
研究題目 Title	NMF とバギングを用いたカルシウムイメージングデータのクラスタリング方法の一検討				

研究背景

脳内の情報伝達には複数のニューロンが関わっている。そのため、同時活動するニューロングループの推定が脳の機能解明につながる可能性がある。脳内の多くのニューロンを観察する方法に、蛍光イメージングの1つであるカルシウムイメージングがある。ニューロンで活動電位が発生（発火）すると細胞内の Ca^{2+} 濃度が上昇する。カルシウムイメージングでは、この Ca^{2+} 濃度上昇を蛍光で可視化する。蛍光イメージングを用いる利点として、(1) 高い空間分解能、(2) 広い観察範囲、(3) 遺伝子工学と併用して興奮性／抑制性ニューロンの同定などをした上での観察ができることが挙げられる。一方、時間分解能が電気生理学的手法よりも低いことが蛍光イメージングの欠点である。 Ca^{2+} 濃度の変化はニューロンの電気的变化よりも遅く、また、カルシウム感受性蛍光分子のキネティクスも影響する。さらに、カメラやレーザースキャンでのサンプリングレートは高くても 100[Hz] 程度であり、ニューロンの個々の発火を全て捉えるには不十分である。

本研究で扱うデータは、睡眠時と覚醒時のニューロンの活動を解析するために計測されたマウスのカルシウムイメージングデータである。8[Hz] のサンプリングレートで1時間おきに15分間のイメージングが計6回行われ、100～200個のニューロンが観測された。8[Hz] のサンプリングレートではニューロンのシナプス伝達を抽出することはできないため、ニューロンの機能的なコネクティビティを抽出することを目標とする。機能的なコネクティビティとは、つまり同時に活動するニューロンが持つ結合のことである。よって本研究では、ニューロンのクラスタリングの問題を解く。クラスタリングではグループに所属しないニューロンの推定とハードクラスタリングを目的とする。提案アプローチによってどれほどの情報が抽出できるかを人工データ実験によって確かめる。

提案アプローチ

ニューロンのクラスタリングには相互相関行列をクラスタリングする方法が考えられる。しかし、相互相関行列ではグループに所属しないニューロンを閾値によって決めることは難しい。提案アプローチではニューロン同士が同じグループになる確率値をバギングによって推定することで閾値を決めやすくする。また、提案アプローチでは nonnegative matrix factorization (NMF) を用いてニューロン間の隣接行列を推定する。NMF の基底数の推定は難しいため、複数の基底数の推定結果

を平均する。NMF の推定量を隣接行列とするのは、異なる基底数で推定量のサイズが変化しないためである。

提案アプローチは2段階に分かれる。まず、(1) ニューロン間の隣接行列 $\hat{A}(X, K)$ を推定し、(2) スペクトラルクラスタリングによってクラスタリングを行う。

ニューロン i の観測時系列は $x_i: \in \mathbb{R}_+^J$ である。 $c_k: \in \mathbb{R}_+^J$ ($k = 1, \dots, K$) をグループ k の活動の時系列とすると、仮定より x_i は c_i の重み付き和として表される：

$$x_i = \sum_{k=1}^K d_{ik} c_k + \eta_i. \quad (1)$$

ただし、 $d_{ik} \in \mathbb{R}^+$ で、 $\eta_i: \in \mathbb{R}^J$ はガウスノイズの時系列である。(1) は行列形式で以下のように表現できる：

$$X = DC + H. \quad (2)$$

ただし、 $X \in \mathbb{R}_+^{I \times J}$, $D \in \mathbb{R}_+^{I \times K}$, $C \in \mathbb{R}_+^{K \times J}$, $H \in \mathbb{R}^{I \times J}$ である。また、 D の要素 (i, k) は d_{ik} , C の i 行は c_i , H の i 行は η_i である。

NMF は行列分解手法の一つであり、1つの非負行列を2つの非負行列の積として分解する。提案アプローチでは、NMF を用いて (2) を解いて D と C を求める。

NMF での推定量を $\hat{A}(X, k) \in \{0, 1\}^{I \times I}$ とする。NMF の寄与率行列 $P \in \mathbb{R}^{I \times K}$ の要素を次のように定義する：

$$p_{ik} = \frac{\|d_{ik} c_k\|_1}{\sum_{l=1}^K \|d_{il} c_l\|_1}$$

要素 p_{ik} はニューロン i に対する基底 k の寄与率という意味である。 $\hat{A}(X, k)$ は、 P の行ごとに最大値を1、それ以外を0とした行列 G を用いて、

$$\hat{A}(X, k) = GG^\top,$$

と定義する。 $\hat{A}(X, k)$ は対称行列で、要素 \hat{a}_{ij} はニューロン i とニューロン j が同じ基底に所属していると推定したら1、していないと推定したら0となる。

バギングを用いて隣接行列 $\hat{A}(X, K)$, $K = \{K_{min}, \dots, K_{max}\}$ をブートストラップ推定量 \hat{A} の平均として求める。データ X からブートストラップサンプル $\{X^{*b}; b = 1, \dots, B\}$ を作成し、 $\hat{A}(X^{*b}, k)$ を推定する。NMF では基底数を決めなければならないが、確立された推定方法がなく難しい問題である。基底数を間違えた時のリスクを下げるため、異なる基底数 K

のモデル平均をとる． $\bar{A}(X, \mathcal{K})$ は $\hat{A}(X^{*b}, k)$ の期待値として算出する：

$$\bar{A}(X, \mathcal{K}) = \frac{1}{|\mathcal{K}|} \sum_{k \in \mathcal{K}} \frac{1}{B} \sum_{b=1}^B \hat{A}(X^{*b}, k).$$

$\bar{A}(X, \mathcal{K})$ のクラスタリングにはスペクトラルクラスタリングを用いる．スペクトラルクラスタリングは，隣接行列からグラフラプラシアンを求め，その固有ベクトルを k-means などでもクラスタリングする手法である． $\bar{A}(X, \mathcal{K})$ は隣接行列なのでスペクトラルクラスタリングと相性が良いと考えられる．

実験結果

ニューロン 1000 個の活動のシミュレーションを行い，人工のカルシウムイメージングデータを作成して，提案アプローチがどこまでうまく働くかを検証した．その結果，NMF を使う有用性，バギングとモデル平均を用いることで精度の向上が確認できた．

目次

1	序論	1
1.1	カルシウムイメージングデータ	1
1.2	関連研究	3
1.2.1	脳データの解析	3
1.2.2	カルシウムイメージングの解析例	4
1.3	目的	4
2	準備	7
2.1	NMF	7
2.2	バギング	7
2.3	スペクトラルクラスタリング	7
3	NMF とバギングを用いたクラスタリング方法	9
3.1	隣接行列の推定	9
3.1.1	数理モデル	9
3.1.2	NMF による \hat{A} の推定	10
3.1.3	バギング	12
3.1.4	NMF の基底数	12
3.1.5	モデル平均	13
3.1.6	NMF の一意性	13
3.2	\bar{A} を用いたクラスタリング	14
3.2.1	スペクトラルクラスタリングのクラスタ数の決め方	14
3.2.2	評価	15
4	実験	17
4.1	人工データ実験	17
4.1.1	シミュレーション	17
4.1.2	\bar{A} の推定に関する実験	21
4.1.3	\bar{A} のクラスタリングに関する実験	23
4.2	実データ解析	25
4.2.1	実データ	26
4.2.2	実験設定	26
4.2.3	結果	26
5	結論	29
付録 A	人工データの作成方法	37
付録 B	考案した NMF の制約	39
1	分解能の決定	39
2	スケール除去	39
3	重複除去	40
4	時間方向への制約	41

付録 C	NMF のモデルエビデンス	43
0.1	ラプラス近似	43
0.2	ブートストラップによる近似	44
0.3	実験	44

第1章 序論

脳の機能解明のためにニューロンの活動解析が行われている。ニューロンは複数で活動することによって情報処理を行っているため、集団に注目して解析することが重要である。ニューロンの観察方法として、パッチクランプ法、細胞内記録法、細胞外記録法などの電気生理学的な手法が挙げられる。これらの手法は十分な時間分解能かつ細胞レベルでニューロンを観察することができる。しかし、電気生理学的な手法では電極から近いニューロンしか電位変化の計測できないため、観察できるニューロンの数は数十から多くても数百程度である。

より多くのニューロンを観察するために、蛍光イメージングの1つであるカルシウムイメージングという手法が用いられる。ニューロンで活動電位が発生（発火）すると細胞内の Ca^{2+} 濃度が上昇する。カルシウムイメージングでは、この Ca^{2+} 濃度上昇を蛍光で可視化する。具体的には、 Ca^{2+} と結合すると蛍光強度が変化する蛍光分子を細胞内に発現させておき、 Ca^{2+} 濃度を蛍光強度として蛍光顕微鏡で観察する。蛍光イメージングを用いる利点として、(1) 高い空間分解能、(2) 広い観察範囲、(3) 遺伝子工学と併用して興奮性／抑制性ニューロンの同定などをした上での観察ができることが挙げられる。一方、時間分解能が電気生理学的手法よりも低いことが蛍光イメージングの欠点である。 Ca^{2+} 濃度の変化はニューロンの電気的变化よりも遅く、また、カルシウム感受性蛍光分子のキネティクスも影響する。さらに、カメラやレーザースキャンでのサンプリングレートは高くても100[Hz]程度であり、ニューロンの個々の発火を全て捉えるには不十分である。これより、カルシウムイメージングデータからはニューロン活動の相関しか取ることができないと考えられる。

ニューロン集団の解析にはクラスタリングやネットワーク構造推定などが考えられる。相関しか取れないカルシウムイメージングデータからは向きのある有向グラフの推定は難しく、向きのない無向グラフの推定しかできない。クラスタリングは、同じグループ内のニューロン同士がエッジを持つ無向グラフとしても考えられる。結果の解釈はクラスタリングの方がしやすい。ニューロンの集団解析で考えられる最も単純な方法は、相互相関行列を用いてクラスタリングを行う方法である。しかし、ニューロンが同時活動する時間が短い場合には同じグループ内のニューロンの相関がグループ外のニューロンと比べて高くならず、意味のある情報が抽出できないと思われる。また、どのグループにも所属しないニューロンの存在も考えられるが、相互相関行列ではこのようなニューロンの判定は難しい。

本研究では、低い時間分解能のカルシウムイメージングデータからニューロンをクラスタリングするアプローチを提案し、人工データ実験を通してどの程度の情報が抽出できるかを確認する。提案アプローチでは、ニューロンが同じグループに所属する確率値を無向グラフの隣接行列として推定し、それをスペクトラルクラスタリングによってクラスタリングする。グループに所属しないニューロンは、他のニューロンとグループになる確率が低いものとして推定することができる。隣接行列の推定には数値モデルに基づいて non-negative matrix factorization (NMF) を用いることで、相互相関行列や他の行列分解手法よりも精度の良いクラスタリングができると考えられる。

1.1 カルシウムイメージングデータ

本節では、カルシウムイメージングデータがどのようなデータなのかを説明し、どの程度の情報が取り出せるかを議論する。

ニューロンは1[ms]単位で活動電位が発生する(発火)。活動電位は細胞内外のイオン濃度

が局所的に変化することによって生じる。活動電位は細胞体から軸索を伝わり、シナプスを介して結合している別のニューロンに伝わる。哺乳類の皮質ニューロンにおいて、ニューロンからニューロンへシナプスを介して活動電位が伝わるには数十 [ms] かかる [1]。このように活動電位を伝えることによってニューロン間で情報がやり取りされる。

ニューロンが他のニューロンとコネクションを持つ状態のことをコネクティビティという。脳のコネクティビティには、anatomical connectivity と functional connectivity と effective connectivity の3種類がある [2]。Anatomical connectivity はニューロンが物理的に繋がっていることを指す。Functional connectivity を持つ2つのニューロンは活動に相関があり、機能が同じことを指す。例えば、ある物を見た時、それに特異的に反応するニューロン群は functional connectivity を持つ。Effective connectivity は、1つのニューロンから別のニューロンへ情報が伝わることを指す。脳の機能解明のためには functional connectivity と effective connectivity の推定が必要である。脳内にはセルアセンブリというニューロン群が存在すると言われており [3]、セルアセンブリ内のニューロンはある刺激の元で同時に活動することが知られている。セルアセンブリは functional connectivity を持っていると考えられるので、functional connectivity を推定することはセルアセンブリの推定につながる。Effective connectivity は情報伝達の順番が分かるため、これも脳の機能解明につながる。例えば、リアクチベーション現象があることが知られている。リアクチベーション現象とは、ニューロンが同じ順番で発火することが繰り返されることであり、記憶の定着などにも関わっていると言われている。データによってどのコネクティビティの情報を取り出せるかは異なり、電気生理では effective connectivity が推定できるが、サンプリングレートの低いカルシウムイメージングデータでは functional connectivity までしか推定ができない。

1個1個のニューロンを1[ms]単位で計測できればニューロン全ての活動を計測できるが、そのような技術は存在しない。ニューロンの計測方法には様々なものがあり、それぞれ計測可能な時間分解能と空間分解能が異なる。計測方法別の分解能については [4] の Fig 1 が分かりやすい。EEG, PET, fMRI は脳の一部のニューロンの活動によって生じた電位、血流、代謝量の変化を計測する。これらの手法ではニューロン単位の計測は行えないが、脳全体を計測することができる。電気生理学的な手法やカルシウムイメージングではニューロン単位で膜電位の変化やカルシウムイオン濃度の変化を計測する。これらの手法ではニューロン1個1個を計測できるが脳全体を計測することはできない。

カルシウムイメージングとはニューロン内のカルシウムイオン濃度を可視化することでニューロンの活動を計測する手法である。観察したい個体のニューロンにカルシウムイオンと結合する蛍光タンパク質を発現させると、細胞内のカルシウムイオン濃度に応じて蛍光強度が変化する。ニューロンが発火するとカルシウムイオンが流入するため蛍光強度が上昇し、その後徐々に蛍光強度は減少する。ニューロンを蛍光顕微鏡で観察することで図 1.1 のような蛍光強度を反映した画像が得られる。その画像から個々のニューロンの蛍光強度のデータを取得できる。

電気生理学的な手法と比べた時のカルシウムイメージングの利点として、ニューロンの位置情報が分かるため同じニューロンを複数回観察できることとより多くのニューロンの測定ができることが挙げられる。また、ニューロンには興奮性ニューロンと抑制性ニューロンの二種類があり、カルシウムイメージングではこの種類を見分けることができる。二種類の蛍光タンパク質を用いて、片方の蛍光タンパク質を抑制性ニューロンだけに発現させることで興奮性・抑制性が分かる。

カルシウムイメージングが電気生理学的な手法より劣る点として、時間分解能が挙げられる。カルシウムイメージングのサンプリングレートは測定機器に限界があり [5]、通常 1-50[Hz] 程度である。ニューロンの発火は約 1[ms] で、シナプスを介して発火が伝わるのは 40[ms] 以内 [6] なので、ニューロンの発火を個別に観測することはできず、低いサンプリングレートでは発火の伝達も捉えることができない。また、カルシウムイオン濃度の変化は発火のタイムスケールより長い。発火後に蛍光強度が変化し始めるのは数 [ms] 後、蛍光強度がピークに達するのは数 100[ms] 後である。一方、蛍光強度が元の値に戻るまでには、数 100[ms] か

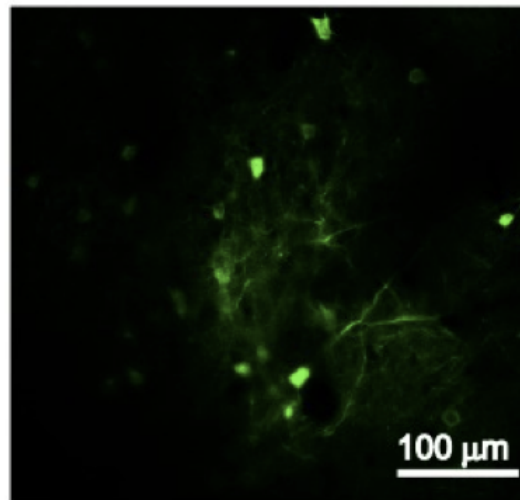


図 1.1: カルシウムイメージングで観察できる画像の例.

ら数 1000[ms] かかる [7]. また, 蛍光タンパク質の性能によっても時間遅れが生じる.

本研究で扱うデータのサンプリングレートは 8[Hz] で, シナプス伝達一つを見るには不十分である. 顕微鏡で観測すると活動電位が伝わる順番は分からなくなるため, カルシウムイメージングデータからはニューロンの活動の相関の情報しか得ることができない. これより, このデータではニューロン間のシナプス伝達を推定するのではなく, 機能的に同じニューロン, つまり同時に活動するニューロングループを推定する問題が適していると考えられる. 機能的に同じニューロンは, 3つの脳のコネクティビティのうち functional connectivity で繋がっているニューロンを指す.

1.2 関連研究

1.2.1 脳データの解析

カルシウムイメージングは時間解像度が低く空間解像度が高いデータが取得できる手法である. 同じ特徴である fMRI に対するデータの解析手法を紹介する.

fMRI データの解析は model based な手法と data-driven な手法がある [8]. Model based な手法は cross-correlation analysis (CCA) や coherence analysis (CA) などが挙げられ, 各指標に基づいて脳部位 (ROI) の機能的な結合 (前述の functional connectivity) を推定する. CCA では 2つの ROI の相関を使う. CA では, 2つの ROI の周波数ドメインでの相関を見る. fMRI の BOLD 信号の周期は 10[s] ほどなので, 0.1[Hz] 以下の周波数ドメインでの相関を見る. 結合の有無は各指標の閾値で決めなければならない. Data-driven な手法は更に行列分解とクラスタリングに分けることができる. 行列分解には principal component analysis (PCA) や singular value decomposition (SVD), independent component analysis (ICA) などが挙げられ, BOLD 信号を複数の成分に分解することができる. 同じ成分の重みが多い ROI は結合があるとする. ICA は成分数を事前に与える必要がある. クラスタリングは fuzzy clustering analysis (FCA) や hierarchical clustering analysis (HCA) などがある. クラスタリングではデータ間の距離を定義する必要がある. FCA はある ROI が複数のクラスタに存在できるクラスタリング手法で, クラスタ数を事前に決める必要がある. HCA はクラスタ数を決めずに ROI やクラスタが近いか遠いかを見ることができる.

本研究では, 行列分解手法の 1つである NMF を用いて隣接行列を推定する. 数理モデルを仮定すると, カルシウムイメージングデータには ICA や PCA より NMF の方がデータに

適している。

1.2.2 カルシウムイメージングの解析例

カルシウムイメージングデータの解析には二種類考えられる。まずは、カルシウムイメージングデータからスパイクを推定し、そのスパイク列を解析する方法である。Vogelstein らは逐次モンテカルロ法を用いてカルシウムイメージングデータからスパイク推定を行なった [9]。しかし、カルシウムイメージングデータでも低いサンプリングレートで計測されたデータではこの方法は使えない。

もう一つは、データから直接ニューロンの活動を解析する方法である。Mishchencko はベイズ推定を用いたニューロンの結合推定を行なった [10]。しかし、カルシウムイメージングのサンプリングレートが 30Hz 以上でないと意味のある結合推定結果は得られないと報告している。また、Stetter らは transfer entropy (TE) を用いて培養された興奮性ニューロンの結合推定を行なった [11]。この手法では、2つのニューロン間の TE を計算し、向きのあるエッジを推定している。また、ニューロンのネットワーク構造を仮定した人工のカルシウムイメージングデータを作成し、推定精度を議論している。しかし、本研究で扱うデータはサンプリングレートが低いため、向きの情報は取り出せない。Ikegaya らは蛍光強度データの一次微分を閾値で切ることによって発火のタイミングの情報を取り出し、テンプレートマッチングによってリアクチベーション現象を解析した [12]。この研究では 1Hz のデータからもリアクチベーション現象を確認できていた。しかし、本研究ではニューロン集団を推定したいため、この方法はあまり適さない。

Molter らはカルシウムイメージングデータからニューロングループを抽出する方法を 8 つ人工データ実験と共に試した [13]。手法は大きく 2 つに分けられ、ニューロンペアの相関を見るものと、全てのニューロン活動の状態を見るものである。前者では固有値分解によって相関行列を作成した後、ICA や Promax rotation によってグルーピングを行う。後者では、ニューロンの活動から SVD, k-means クラスタリング、スペクトラルクラスタリングなどを用いてグルーピングを行う。後者の方法では、各グループの時間方向の活動を平均をとるなどして、グループの活動としていた。人工データは、ニューロンをポアソン分布にしたがって発火させ、発火からカルシウムイメージングの観測データに変換していた。同じグループに所属するニューロンは発火確率を同じ時間帯にあげることで表現していた。用いられていた手法では、ニューロンが複数グループに所属する場合やグループに所属しない場合も対応ができる。最も良いと結論づけられていた手法は ICA-CS と SGC という手法だったが、どちらの手法も閾値などのパラメータを決めなければならない。

Ghandor らは学習に関係するニューロン (engram cell) 群について NMF を用いて解析を行った [14]。実験は複数セッションに分けて行われており、それぞれのセッションで NMF を用いてニューロングループとその活動に分解していた。基底数は AICc で決めていた。セッションごとに推定されたグループが近いかを cosine similarity で計った結果、engram cell では non-engram cell よりも繰り返し活動するグループが多かった。NMF によってニューロングループがどれほど抽出できるかは論じられていなかった。

1.3 目的

カルシウムイメージングデータは上述のように多くのニューロンを観測できる利点がある一方、時間分解能が低いという欠点を持つ。本研究で扱うデータは低いサンプリングレートで観測されたデータである。ニューロンは複数で活動することで情報伝達を行っており、睡眠時に多数のニューロンが同時に活動する現象も確認されている。これより、本データでは同時に活動するニューロングループを推定して解析を行うことが適当である。

本研究の目的は、カルシウムイメージングデータから同時に活動するニューロングループを推定することである。紹介した関連研究は、本研究で扱いたいデータのサンプリングレ

トでは使えないものや、パラメータを適切に決めなければならないものだった。提案アプローチでは、数理モデルに基づいてNMFを用いてニューロンの隣接行列を推定し、スペクトラルクラスタリングを用いて隣接行列からクラスタリングを行う。隣接行列はバギングを用いて、2つのニューロンが同じグループになる確率値をつける。そのため、グループに所属しないニューロンの推定や、どれくらい結果を信用して良いのかがわかる。その際も閾値を決めなければいけないが、確率値という分かりやすい指標のため、既存手法より簡単だと思われる。NMF自体も基底数を事前に決めなければいけないが、大体の基底数を決めればその基底数周りの結果を平均することでパラメータによる結果の劇的な変化を抑えることができる。また、人工データ実験によって提案アプローチがカルシウムイメージングデータからニューロングループの情報をどの程度取り出せるかを確かめる。

本論文の構成は以下の通りである。第2章では数理モデルを元にしたクラスタリングのアプローチを説明する。第3章では人工データの作成方法と人工データ実験の結果を述べる。第4章では実データ解析の結果を述べる。第5章では採用に至らなかった検討事項について述べ、第6章に本研究をまとめる。

第2章 準備

本節では，提案アプローチで用いる NMF，バギング，スペクトラルクラスタリングについて説明する．

2.1 NMF

2.2 バギング

2.3 スペクトラルクラスタリング

第3章 NMFとバギングを用いたクラスタリング方法

本章では、ニューロン同士が同じグループである確率を推定し、その情報を元にクラスタリングを行う方法について説明する。

3.1 隣接行列の推定

3.1.1 数理モデル

カルシウムイメージングデータに対していくつかの仮定をおいた。

仮定 1

グループが K 個存在し、同じグループ内のニューロンは同じ時刻に活動する。ニューロンは複数のグループに所属することができる。観測時間内ではグループに属するニューロンは変化しない。

仮定 2

複数のグループが同時に活動する時、属するニューロンは被らない（ニューロンが属するグループは同時には活動しない）。

序論で述べたように、脳内にはセルアセンブリが存在すると言われているため、グループが K 個存在するという仮定は妥当である。また、シナプス結合は短期間で変化するものではないため、仮定 1 は妥当である。仮定 2 の概念図を図 3.1 に示す。仮定 2 は、主に線形モデルで扱うための仮定であり後述する。脳内のニューロンの働きを考えても、ニューロンはある時刻で 1 つの機能しか持たないと考えられるため、所属する複数グループが同時に活動することは考えにくい。逆に、同時に活動する複数グループは、同じ機能を持った 1 つのグループであると考えられる。

これらの仮説をもとに、数理モデルを構築する。観測データ X は $X \in \mathbb{R}_+^{I \times J}$ とする。ただし、 \mathbb{R}_+ を非負の実数の集合、 J を観測時系列の長さとする。ある行列 A の i 行を $a_{i:}$ 、 j 列を $a_{:j}$ 、 (i, j) 要素を a_{ij} または $[A]_{ij}$ と表記する。よって、ニューロン i の観測時系列は $x_{i:} \in \mathbb{R}_+^J$ である。

行ベクトル $c_{k:} \in \mathbb{R}_+^J$ ($k = 1, \dots, K$) をグループ k の活動の時系列とすると、仮定より $x_{i:}$ は $c_{k:}$ の重み付き和として表される：

$$x_{i:} = \sum_{k=1}^K d_{ik} c_{k:} + \eta_{i:}. \quad (3.1)$$

ただし、 $d_{ik} \in \mathbb{R}_+$ で、 $\eta_{i:} \in \mathbb{R}^J$ は i.i.d. なガウスノイズの時系列である。カルシウムイメージングのノイズはポアソン分布に従う光子ノイズであるが、光子数が多い場合はガウス分布で近似できる [15]。仮定 2 を置くことによって、式 (3.1) の線形モデルを考えることができる。仮定 2 が無い場合、あるニューロンの蛍光強度は複数グループからの影響によって上限なく上昇できてしまう。実際は、ニューロンの蛍光強度の最大値は蛍光タンパク質の量で決まるため、蛍光強度は最大値で飽和する。その場合式 (3.1) には飽和を組み込まなければい

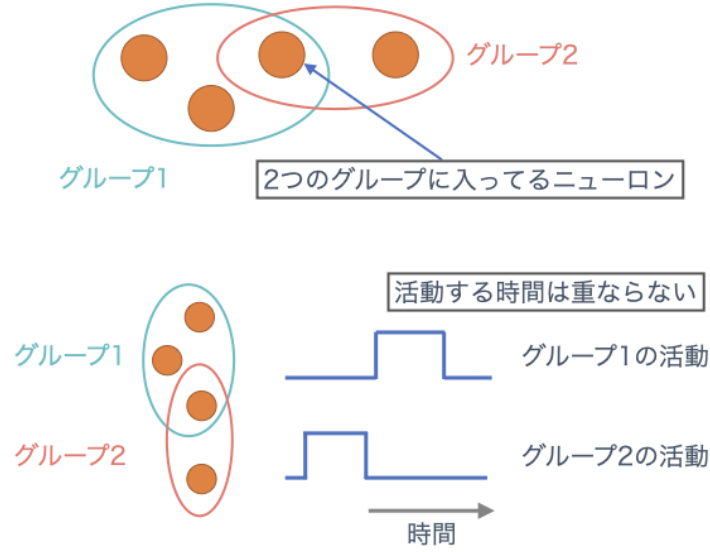


図 3.1: 仮定 2 の概念図. 所属するニューロンに被りがある 2 つのグループは活動時間が被らない.

けない. 本論文では, 複数グループが同時に活動する場合はそれらも同じグループとみなし, 単純な線形モデルで表す.

式 (3.1) は行列形式で以下のように表現できる:

$$X = DC + H. \quad (3.2)$$

ただし, $D \in \mathbb{R}_+^{I \times K}$, $C \in \mathbb{R}_+^{K \times J}$, $H \in \mathbb{R}^{I \times J}$ である. また, D の要素 (i, k) は d_{ik} , C の i 行は $c_{i:}$, H の i 行は η_i である. NMF のモデルを図 3.2 に示す.

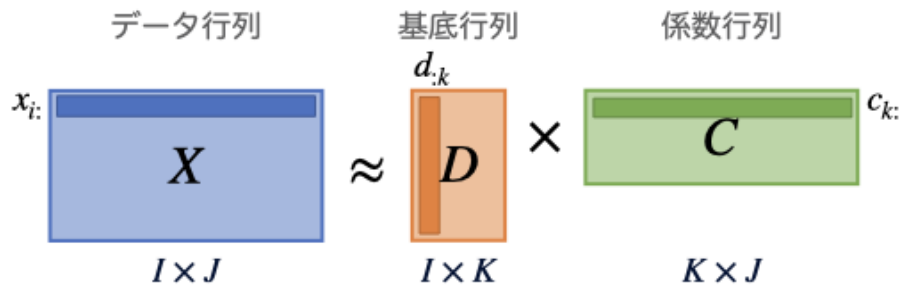


図 3.2: NMF のモデル.

3.1.2 NMF による \hat{A} の推定

非負行列因子分解 (nonnegative matrix factorization; NMF) [16] は行列分解の手法の一つである. NMF は非負行列を 2 つの非負行列の積として分解する:

$$X \approx DC.$$

X , D , C は上述の行列であり, 式 (3.2) は NMF の問題として解くことができる. D と C を求めるために NMF では以下の最適問題を解く:

$$\arg \min_{D \geq 0, C \geq 0} \|X - DC\|_F^2.$$

基底数 K は事前に決めなければならないパラメータである.

NMF の寄与率行列 $P \in [0, 1]^{I \times K}$ の要素を次のように定義する:

$$p_{ik} = \frac{\|d_{ik}c_{k:}\|_1}{\sum_{l=1}^K \|d_{il}c_{l:}\|_1}.$$

また, 概念図を図 3.3 に示す. ニューロン i の観測系列 $x_{i:}$ は $d_{ik}c_{k:}$ ($k = 1, \dots, K$) で再構成

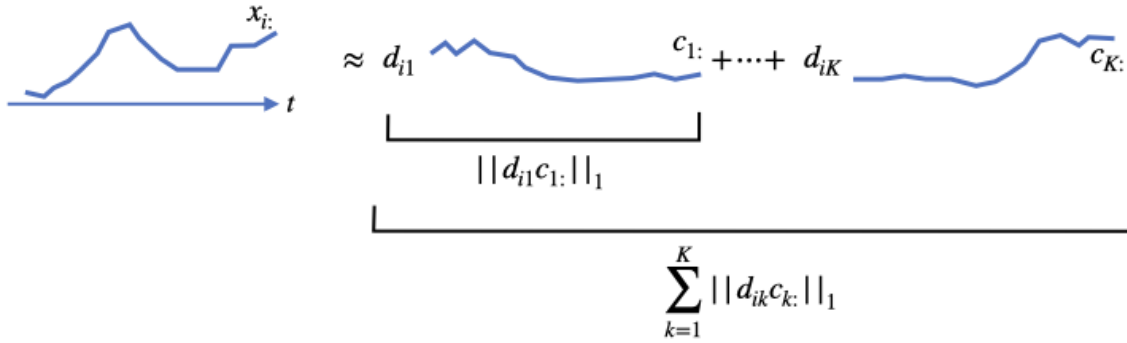


図 3.3: 寄与率の概念図.

されている. 観測系列は非負なので, $\|d_{ik}c_{k:}\|_1$ は基底 k の情報が $x_{i:}$ の再構成のうちどれくらいを占めているかを示す. よって, 要素 p_{ik} は, ニューロン i の観測系列 $x_{i:}$ の再構成にどれくらい基底 k が寄与したかという量である. NMF の目的関数に C の行和が 1 という制約を入れると $P = D$ となる.

NMF での推定量 $\hat{A}(X, K) \in \{0, 1\}^{I \times I}$ とする. $\hat{A}(X, K)$ は寄与率行列 P から作成する. まず, P と同じサイズの行列 $G \in \{0, 1\}^{I \times K}$ を作る. G は, P の各行について最大値のみを 1, それ以外を 0 とした行列である.

$$g_{ij} = \begin{cases} 1 & (j = \arg \max_j p_{ij}) \\ 0 & (\text{otherwise}) \end{cases}$$

推定量 $\hat{A}(X, K)$ を

$$\hat{A}(X, K) = GG^\top,$$

と定義する. これより \hat{A} は対称行列で, 要素 \hat{a}_{ij} はニューロン i とニューロン j が同じ基底に所属していたら 1, していなかったら 0 となる. この推定量は cluster ensemble でも用いられている pairwise similarity[17] と似たものになっている.

この推定量を使う利点は 2 つある. 1 つ目は, NMF では D の行と C の列は毎回同じ順番ではないため, その対応を考えなくて良い点である. 2 つ目は, 基底数の情報がなくなるため, 異なる基底数の結果と比較できるところである. 2 つ目の利点があるからこそ後述のモデル平均を行うことができる.

\hat{A} のみでもクラスタリングは行える. しかし, この $\{0, 1\}$ の行列には「本当は 0 だがたまたま 1 になった要素」やその逆が存在する. 全てのニューロンを 1 つの基底に割り振っているため, 本来どのグループにも所属しないニューロンの判定もできない. そこで, \hat{A} の要素に信頼度をつけることを次節で考える.

3.1.3 バギング

アンサンブル学習の一つにバギング [18] がある。バギングはある予測器にブートストラップサンプルを入力して、その結果の平均を用いることで精度を向上させる方法である。ブートストラップ法 [19] とは、推定量の分布を近似する方法である。ブートストラップ法によって X が与えられた時の $\hat{A}(X, K)$ の分布を近似することができる。

提案アプローチではブートストラップサンプル $\{X^{*b}; b = 1, \dots, B\}$ から NMF を用いて $\hat{A}^{*b} = \hat{A}(X^{*b}, K)$ を推定する。これより \hat{A}^* の分布が求まるので、 \hat{a}_{ij}^* が 1 となる確率が以下のようにして求まる：

$$Pr(\hat{a}_{ij}^* = 1) = E[\hat{a}_{ij}^* | X].$$

ブートストラップサンプルの作成にはいくつかの方法が考えられる。簡単な 3 種類の方法について説明する。列のサンプリングによるブートストラップは、データ行列 X の各列がデータサンプルとしてみなせるので、列をサンプリングして X^{*b} を作る方法である。ブロックブートストラップでは、複数列を塊としてサンプリングする。この方法は時系列データでのブートストラップで用いられており、時間方向に制約の入った NMF などでは有効である。残差型ブートストラップは一回 NMF のモデルを推定し、推定後の残差をサンプリングして推定した DC に足す方法である。

$$X^{*b} = \hat{D}\hat{C} + H^{*b},$$

ただし、 \hat{D} と \hat{C} は最尤推定量で、 H^{*b} は $X - \hat{D}\hat{C}$ をリサンプリングした行列である。ニューロンごとにノイズの大きさが異なることが想定される場合は、行ごとにリサンプリングを行うのが適切だと考えられる。NMF のモデル式 (3.2) では H は i.i.d. なノイズという仮定を置いているので、残差型ブートストラップが一番モデルに沿ったブートストラップ方法と言える。

3.1.4 NMF の基底数

NMF の基底数を決めるのは難しい問題である。本節では基底数の決め方について述べるが、最適な方法は確立されていない。そこで次節では複数の基底数の結果のモデル平均をとることで、基底数を間違えるリスクを軽減することを考える。

NMF の基底数の決め方にはいくつかのアプローチがある。まずは、専門家や解析者の知識に基づいて決めることである。この方法はデータに対して十分な知識がない時には使えない。

次に BIC [20] や AIC [21] を用いる方法である。これらは漸近理論に基づいた近似を行った情報量基準である。NMF はデータが増えるほどパラメータ数 $(I + J) * K$ が増えるという特徴があり、これらを用いるのは本来不適切である。

次に R の NMF パッケージにも組み込まれている Brunet ら [22] の方法を紹介する。彼らは NMF の推定結果からノード同士が同じ基底に所属するかしないかを表す connectivity matrix $M \in \{0, 1\}^{I \times I}$ を作成する。本論文の推定量 A と同じ意味の行列である。初期値を 20-100 回変化させて M の平均 $\bar{M} \in [0, 1]^{I \times I}$ を計算する。彼らは真の基底数ではこの推定量が 0 か 1 に寄るようになると仮定して、最も \bar{M} が安定する基底数を求める。安定度は $1 - \bar{M}$ とその cophenetic correlation coefficient の Pearson correlation から計算する。

Ubaru ら [23] はブートストラップを用いて NMF を行っており、 D がそれほど変わらない基底数を採用している。推定した D^{*b} 同士の相互相関行列について dissimilarity [24] を測りその平均が最小となる基底数を用いる。

Hutchins らはテストデータに対する RSS (residual sum of squares) が真の基底数以降になるとあまり下がらなくなると論じている [25]。

Bayesian NMF ではギブスサンプリングなどを用いてモデルエビデンスを計算している [26]。

上記で述べた方法の他にも様々な方法が考案されているが、全てのデータに当てはめられるような枠組みは存在しない。

3.1.5 モデル平均

NMF の基底数を決めるのは難しいので、基底数を間違えてもリスクを減らす方法としてモデル平均を用いる。モデル平均とは異なるモデルの推定量の平均をとって精度向上を測る方法である。 \bar{A} は基底数によって行列サイズが変化しないので、異なる基底数の推定結果の平均をとることができる。その場合の推定量は：

$$\bar{A}(X, \mathcal{K}) = \frac{1}{|\mathcal{K}|} \sum_{k \in \mathcal{K}} \frac{1}{B} \sum_{b=1}^B \hat{A}(X^{*b}, k),$$

$$\mathcal{K} = \{K_{min}, \dots, K_{max}\},$$

ただし、平均する最小の基底数を K_{min} 、最大の基底数を K_{max} とする。

NMF の基底数を求めるのは難しい問題なので、モデル平均によって、異なる基底数の推定結果を平均して精度をあまり落とさないようにする。今回扱うデータから推定される \hat{A} は基底数が異なる時に大きくは変化しないと考えられる。

アンサンブル学習では、それぞれのモデルにある程度の推定精度があり [27]、答えに多様性がある方が精度が上がる [28] と言われている。本論文での使い方は異なり、真の基底数が分からない時に推定結果をなまして推定を間違えるリスクを下げるという使い方である。

3.1.6 NMF の一意性

NMF の推定には一意性がなく、ある正則行列 Q を考えた時、

$$\begin{aligned} X &= DC \\ &= DQRC \\ &= D'C', \\ R &= Q^{-1}, \\ D' &= DQ, \\ C' &= RC, \end{aligned} \tag{3.3}$$

のように別の D' と C' が推定される可能性がある。

NMF に一意性がある条件は [29] などでもとめられている。しかし、一意性を持たせるにはかなり条件が狭められる。

NMF と同じように非負行列を分解する手法に nonnegative rank factorization (NRF) [30] がある。NRF ではデータ行列 X を $X = DC$ に分解できる最小の基底数を nonnegative rank $\text{rank}_+(X)$ と定義し、 $\text{rank}(X) = \text{rank}(X)_+$ となる行列 X を対象とする。全ての非負行列は NMF できるが、NRF ができるとは限らない。NRF の計算は NP 困難であるが、[30] では NRF の計算と、NRF を持たない行列に対して MNRS という分解の計算方法を提示している。しかし、 X にノイズが乗っている場合に NRF が存在する条件は調査されていない。

式 (3.3) の場合の D' と C' を用いて作られる寄与率行列 P' と元の寄与率行列 P の関係を考える。簡単のため、 X に行和 1 の正規化を加え C に行和 1 の制約を加える。この時 D の行和も 1 となり、 $P = D$ となる。式 (3.3) より、 $P' = PQ$ となる。この時、 G の作り方には一意がないため推定量 A にも一意がない。

$K = 3$ の場合の図を用いて説明をする。図 3.4 に C の空間内の x_i と D の空間内の d_i を表す。 d_i の和は 1 であるが、軸は図 3.5 のように動くことができる。 G を作成した際にニューロンがどの基底に所属するかを図 3.6 に示した。軸の変化によってニューロンが所属

する基底が変化することがある。その結果、推定量 \hat{A} も変化するため、NMF は \hat{A} の推定に関して一意性はない。第 3 章の実験を通して一意性がなくても精度に向上は見られることを確かめる。

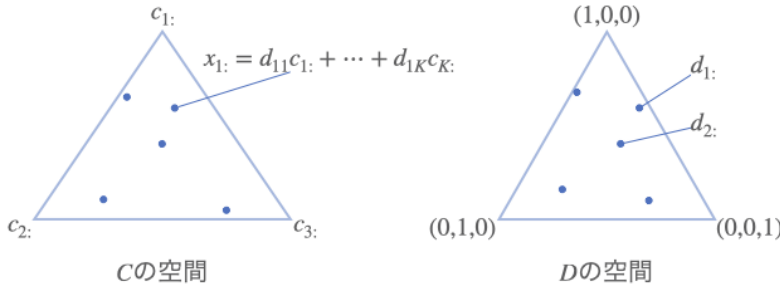


図 3.4: C の空間と D の空間。

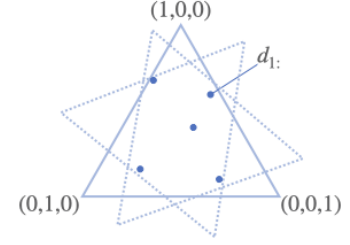


図 3.5: 変換 Q によって D の空間の軸は回転・移動する。

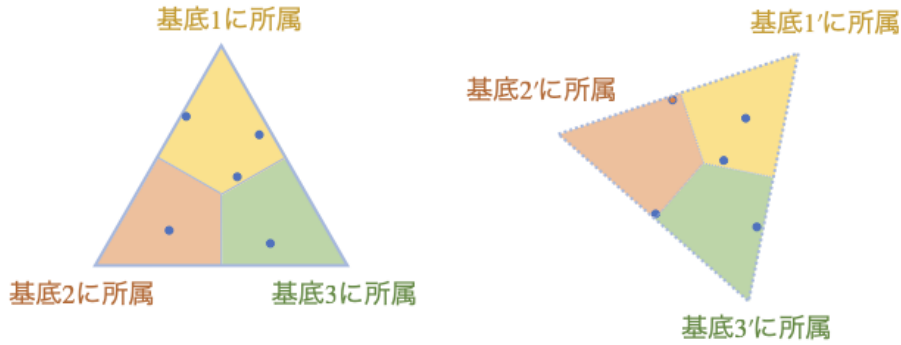


図 3.6: D の空間の軸の変化によってニューロンが所属する基底が変化する。

3.2 \bar{A} を用いたクラスタリング

\bar{A} のクラスタリングにはスペクトラルクラスタリングを用いる。スペクトラルクラスタリングは、隣接行列からグラフラプシアンを求め、その固有ベクトルを k-means などでクラスタリングする。他のクラスタリング手法も使えるが、 \bar{A} をグラフとして見るができるため、スペクトラルクラスタリングを用いる。

3.2.1 スペクトラルクラスタリングのクラスタ数の決め方

クラスタ数の決め方には様々な方法がある。クラスタリング手法によらない方法には、安定性を見る方法 [31] や Gap 統計量 [32] がある。スペクトラルクラスタリングに特化した方法としては、固有値ギャップを見る方法 [33] がある。固有値ギャップは固有値を小さい順にプロットした際に、一気に固有値が大きくなる箇所であり、その順位をクラスタ数とする。

クラスタ数が明瞭な類似度行列について固有値ギャップと Gap 統計量を試した結果、固有値ギャップの方が安定してクラスタ数を決定できたため、固有値ギャップを用いる。固有値ギャップは目で見て判断しなければならないが、本実験では固有値の差分を取って大津の二値化にかけ、差分の大きいグループの最大固有値の順位を固有値ギャップとして計算する。

3.2.2 評価

クラスタリングの評価方法には [13] で用いられていた Best Match score を用いる．二つのクラスタリング結果 $\mathcal{B} = \{B_1, \dots, B_{|\mathcal{B}|}\}$ と $\mathcal{B}' = \{B'_1, \dots, B'_{|\mathcal{B}'|}\}$ があった場合を考える．ただし, B_k はクラスタ k に所属するニューロンの集合である．Best Match score は

$$\text{Best Match score} = \frac{1}{|\mathcal{B}| + |\mathcal{B}'|} \left(\sum_{B \in \mathcal{B}} \max_{B' \in \mathcal{B}'} \frac{|B \cap B'|}{|B \cup B'|} + \sum_{B' \in \mathcal{B}'} \max_{B \in \mathcal{B}} \frac{|B \cap B'|}{|B \cup B'|} \right),$$

である．これは Jaccard 係数のクラスタ数の平均である．2つのクラスタリング結果が近いほど Best Match score は高くなる．評価には \mathcal{B} をクラスタリング結果, \mathcal{B}' を真のクラスタ集合として計算する．

第4章 実験

本章では、人工データ実験を通して提案アプローチの性質調査と実データ解析の結果を述べる。

4.1 人工データ実験

4.1.1 シミュレーション

ニューロン集団のカルシウムイメージングデータをシミュレーションによって作り、解析手法を評価する。シミュレーションでは1) ニューロンのネットワーク構造を作成し、2) スパイクのシミュレーションを行い、3) 蛍光強度の観測データに変換する。

ネットワーク構造

シミュレーションに用いるニューロンの個数を I として、ニューロンのネットワーク構造を $S \in \{0, 1\}^{I \times I}$ とする。 s_{ij} はニューロン i からニューロン j へ活動電位が伝わるかを表している。本節では S の作り方を説明する。

ニューロンのネットワーク構造には small world network[34] を用いる。Small world network はノード数、張り替え確率、初期次数を決めることによってネットワークを作成するアルゴリズムである。初期次数は、ニューロンが平均何個のニューロンとシナプス結合を持つかという変数である。張り替え確率は、初期次数によって作成された規則的なグラフのエッジをランダムに張り替える確率である。そのため、エッジのうち何割が遠くのニューロンとつながっているかを表す変数である。

実際のニューロンを small world network によって表すために、初期次数と張り替え確率を実データから決める。今回はこの値はニューロンのコネクションの割合と相互のコネクションの割合から決める。興奮性ニューロン同士の6.7%であり、そのうち双方向のコネクションの割合は24%である[35]。発達中マウスの興奮性ニューロンから抑制性ニューロンへのコネクティビティと抑制性ニューロンから興奮性ニューロンへのコネクティビティはどちらも78%であった[36]。成熟したマウスではより少ないと思われるが、データが見つからなかったため、40%とした。相互のコネクションの割合がランダムにエッジを作るよりも高いのは、近いニューロンにコネクションが作られやすいからだと考えられる。これらのデータを実現するように初期次数と張り替え確率を調整した。用いたパラメータを表4.1に示す。抑制性ニューロン同士のコネクティビティは分からないため、興奮性ニューロンと同じにしている。

結合の種類	初期次数	張り替え確率
同種類のニューロン間	$0.0335N$	0.3
興奮性ニューロンと抑制性ニューロン間	$0.2I$	0.3

表 4.1: ネットワーク構造のパラメータ値

実際のネットワーク構造の作り方を説明する。ネットワーク構造は興奮性ニューロン同士の結合、抑制性ニューロン同士の結合、興奮性ニューロンと抑制性ニューロン間の結合の3つに分けて作成する。まず、全ニューロンのうち抑制性ニューロンと興奮性ニューロンのイ

ンデックスを決めておく．全てのニューロンについて表 4.1 に従ってネットワークを作成し，それぞれに対応する隣接行列の上三角または下三角行列を取り出して結合する．作成したいのは向きのある有向グラフなので，上三角行列と下三角行列を分けて作成する．

スパイクシミュレーション

スパイクのシミュレーションに Izhikevich モデル [37] を用いる．このモデルは Hodgkin-Huxley モデルをもとにしており，計算コストが低い．Izhikevich モデルでは，あるニューロンの膜電位が閾値を超えると発火したとみなし，あらかじめ定義したニューロンのネットワーク構造に従って結合を持つニューロンの膜電位を上昇させる．このシミュレーションで設定しなければいけないのは，個々のニューロンの特徴パラメータ，重み付きのネットワーク構造，外部からのランダムな入力である．

まず，個々のニューロンの特徴パラメータについて説明する．このモデルではニューロンごとに4つのパラメータを設定する必要がある，そのパラメータでニューロンを特徴づける．本論文では興奮性ニューロンには regular spiking neurons，抑制性ニューロンには fast spiking neurons を用いる．それらのパラメータを表 4.5 に示す．ただし， r_e と r_i は 0 から 1 の一様分布に従う確率変数である．

ニューロンの種類	a	b	c	d
興奮性ニューロン	0.02	0.2	$-65 + 15r_e^2$	$8 - 6r_e^2$
抑制性ニューロン	$0.02 + 0.08r_i$	$0.25 - 0.05r_i$	-65	2

表 4.2: Izhikevich モデルのパラメータ値

次に，重み付きのネットワーク構造 $W \in \mathbb{R}^{I \times I}$ について説明する．ニューロン i から j へ結合があった場合， w_{ij} はニューロン i が発火した時にニューロン j の膜電位をどれだけ上昇させるかという数値である． W は，前節で作成した S の非ゼロ要素を数値で置き換えることで作成する．その際，同じグループへの興奮性ニューロンからの入力は一様分布 $U(7, 10)$ ，異なるグループの興奮性ニューロンからの結合は標準偏差 $\sigma_w = 1.5$ の対数正規分布の 7 以下の分布，抑制性ニューロンからの結合は一様分布 $U(-10, 0)$ からサンプルする．対数正規分布を用いる理由は [38] のデータに基づく．

$$w_{ij} = \begin{cases} U(7, 10) & (s_{ij} = 1, i \text{ は興奮性ニューロン}, i \text{ と } j \text{ は同じグループ}) \\ \left\{ \frac{1}{\sqrt{2\pi}\sigma_w^2} \exp\left\{-\frac{(\ln w)^2}{2\sigma_w^2}\right\} \mid w \in [0, 7] \right\} & (s_{ij} = 1, i \text{ は興奮性ニューロン}, i \text{ と } j \text{ は異なるグループ}) \\ U(-10, 0) & (s_{ij} = 1, i \text{ は抑制性ニューロン}) \\ 0 & (s_{ij} = 0) \end{cases} \quad (4.1)$$

最後に外部からのランダムな入力について説明する．ニューロンには観測範囲外からの入力がある（以降，外部入力とする）．そのため，シミュレーション中も外部からの電位を乱数としてニューロンの電位に足す．本論文では，ニューロンの活動も外部入力の大きさに表現する．活動していない興奮性ニューロンと抑制性ニューロンにはそれぞれ， $\mathcal{N}(0, 9)$ と $\mathcal{N}(0, 0.01)$ に従う乱数を足す．活動している興奮性ニューロンと抑制性ニューロンにはそれぞれ， $\mathcal{N}(0.5, 9)$ と $\mathcal{N}(0.2, 0.01)$ に従う乱数を足す．これらを表 4.3 に示す．活動していないニューロンへの外部入力は [37] で用いられていたものを採用した．ただし，興奮性ニューロンの活動時の外部入力は変化させた実験もある．

本論文では同時に活動するニューロンを推定するのが目的の 1 つである．ある時間帯にあるニューロングループが活動する時，そのニューロングループには平均値を上げた外部入力を足し，それ以外のニューロンには平均 0 の外部入力を足す．こうすることで，ニューロン

ニューロンの種類	活動時の外部入力	活動していない時の外部入力
興奮性ニューロン	$\mathcal{N}(0.5, 9)$	$\mathcal{N}(0, 9)$
抑制性ニューロン	$\mathcal{N}(0.2, 0.01)$	$\mathcal{N}(0, 0.01)$

表 4.3: シミュレーションに用いる外部入力の値

グループの活動のみ上がる（つまり蛍光強度が上がる）。実際の脳でもこのように外部からの入力によってニューロンの活動を制御していると考えられる。あるニューロングループを活動させる別の方法として、そのグループのハブとなるニューロンにのみ強い外部入力を与える方法も考えられる。

実際にマウスのニューロンの発火頻度がどれくらいなのか [39] を元に表 4.4 に示す。

ニューロンの種類	覚醒時 (Hz)	ノンレム睡眠時 (Hz)	レム睡眠時 (Hz)
興奮性ニューロン	0.76 ± 1.53	0.69 ± 0.86	0.88 ± 1.33
抑制性ニューロン	5.59 ± 7.25	4.69 ± 5.62	4.25 ± 9.43

表 4.4: ニューロンごとの発火頻度の中央値

カルシウムイメージングモデル

スパイクデータからカルシウムイオン濃度を計算する [9] のモデルを用いる：

$$[\text{Ca}^{2+}]_{i,t} - [\text{Ca}^{2+}]_{i,t-1} = -\frac{\Delta}{\tau}([\text{Ca}^{2+}]_{i,t-1} - [\text{Ca}^{2+}]_b) + A n_{i,t} + \sigma_c \sqrt{\Delta} \epsilon_{i,t},$$

ただし、 $[\text{Ca}^{2+}]_{i,t}$ をニューロン i の時刻 t でのカルシウムイオン濃度、 $[\text{Ca}^{2+}]_b$ をカルシウムイオン濃度のベースライン、 Δ を時間幅、 τ は時定数、 A は1つのスパイクでのカルシウムイオン濃度の上がり幅、 $n_{i,t} \in \{0, 1\}$ はニューロン i の時刻 t でのスパイク、 σ_c はノイズの分散、 $\epsilon_{i,t}$ は標準正規分布に従う確率変数である。この人工データでは、蛍光強度がある値から上昇しない飽和状態は考えないこととする。

次に、同論文のモデルを使ってカルシウムイオン濃度 $[\text{Ca}^{2+}]_{i,t}$ をカルシウムイメージングで計測される蛍光強度 $F_{i,t}$ に変換する：

$$F_{i,t} = \alpha [\text{Ca}^{2+}]_{i,t} + \beta + \sigma_F \epsilon_{i,t},$$

α は強度、 β はバイアス、 σ_F はノイズの分散である。表 4.5 に使用したパラメータを示す。何種類かの蛍光タンパク質の性能を調べた論文に [40] がある。この論文の、1秒間に10回発火した時の decay time（蛍光強度が上がり切ってから半分の強度になるまでの時刻）から $\tau = 2.3$ とし、SN 比から $\sigma_c = 0.5$ とした。

観測モデル

実データは 8[Hz] でサンプリングされたデータなので、シミュレーションした蛍光強度を 125[ms] ごとに足し合わせる：

$$x_{i,t'} = \sum_{t=1}^{125} F_{i,t},$$

ここで、 t' はサンプリング後の時刻を表す。

上記の方法で作成した人工データ時系列と実データをそれぞれ図 4.1 と図 4.2 に示す。

パラメータ値	
$[Ca^{2+}]_b$	0.1
Δ	0.001
τ	2.3
A	5.0
σ_c	0.5
α	1.0
β	10
σ_F	1.0

表 4.5: カルシウムイメージングモデルでのパラメータ値

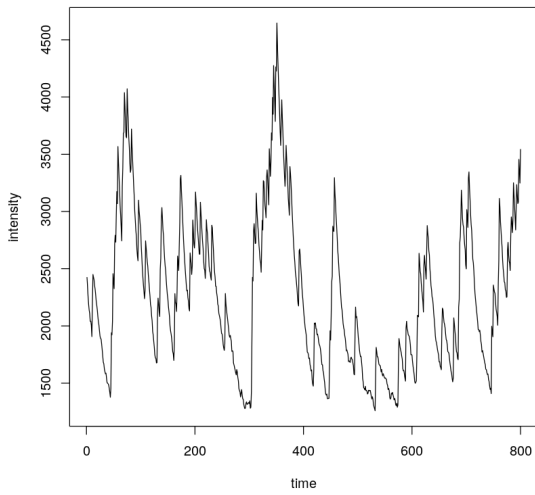


図 4.1: 人工データの時系列.

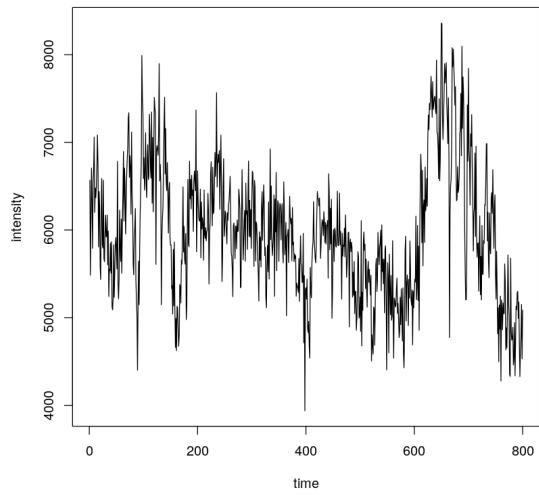


図 4.2: 実データの時系列.

実験設定

800 個の興奮性ニューロンと 200 個の抑制性ニューロンについてネットワーク構造 S を作成し、式 (4.1) に従って重み付きネットワーク構造 W を作成した。 W は全実験を通して固定である。

W を元にカルシウムイメージングのシミュレーションを行った。グループの数は 10 個とした。1つのグループに所属するニューロン数は 50~200 個とした。グループが活動する時間は 5[s] ごとに変えた。ニューロングループは同時に 2 つまで活動することができる。全 1000 個のニューロンのうち、解析に用いるのは固定された 100 個の興奮性ニューロンのみとする。作成するデータの種類の種類は以下の 3 種類とした。

1. ニューロンは 1 つのグループに必ず所属する
2. ニューロンは 1 つのグループに所属するかグループに所属しない
3. ニューロンは 1 つか 2 つのグループに必ず所属し、同じニューロンが所属しているグループ同士の活動は被らない

NMF の基底数は 8 から 12 までをそれぞれ 30 回ブートストラップを行った結果を用いる。NMF の更新則は Nesterov 更新 [41] を用いる。NMF は初期値に依存するため、本章で

は「NMF1回の結果」は「初期値を20回変えてNMFを行い目的関数が最小となった結果」とする。20回は少ないかもしれないが（1000回ほどこの操作を行っている論文もある）、実験サイクルを回すためにこの値にした。

4.1.2 \bar{A} の推定に関する実験

本節では \bar{A} の推定に関する結果について述べる。実験は、5[s]ごとに1つか2つのグループが活動するデータを105[s]シミュレーションさせた結果を用いた。ただし、最初の5[s]はシミュレーション数値の安定性のため解析から除外した。

$\hat{A}(X, K)$ の一意性

前章で述べた通り、NMFには一意性がないため $\hat{A}(X, K)$ にも一意性がない。具体的には、1つの人工データについて初期値のみを変化させてNMFを行うと推定される $\hat{A}(X, K)$ は同じではない。 A の上三角の要素について1と推定された頻度を調べる。収束性と前章で述べたように寄与率が D の空間で取りうる値の範囲を狭めるため、 C の行和を1とする制約を入れている。 X に正規化を加えなかった結果を図4.3に、 X に行和1の正規化を加えた結果を図4.4に示す。 X に行和1の正規化を加えると、 D の自由度は下がる。

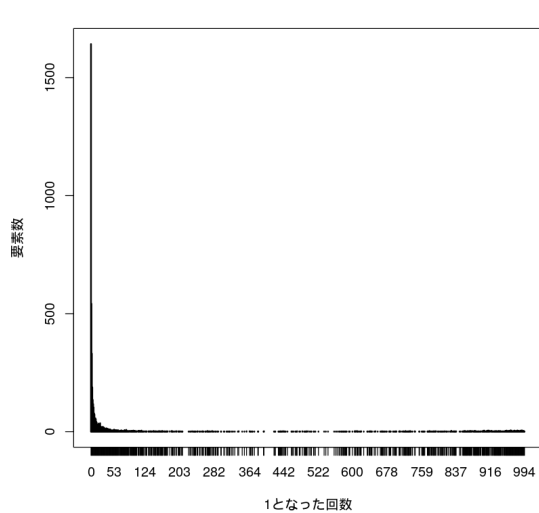


図 4.3: X を正規化せずに初期値を1000回変化させてNMFから $\hat{A}(X, K)$ を推定し、各要素について1と推定された頻度。

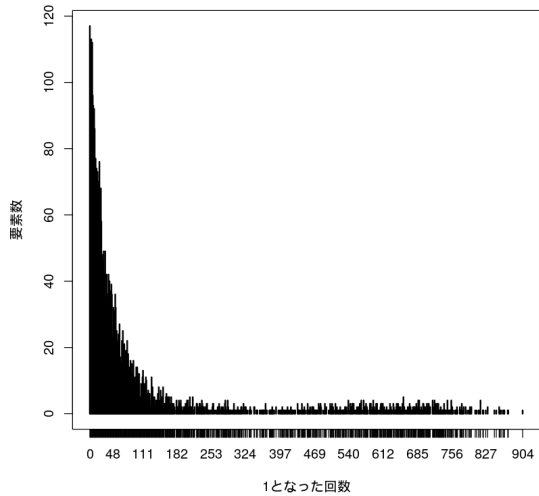


図 4.4: X を行和1で正規化して初期値を1000回変化させてNMFから $\hat{A}(X, K)$ を推定し、各要素について1と推定された頻度。

$\hat{A}(X, K)$ に一意性がある場合は、ある $\hat{A}(X, K)$ の要素が1と推定される回数は0か1000になる。結果より、そうっていないので $\hat{A}(X, K)$ に一意性がないのがわかる。また、図4.4より X を正規化しない方がばらつきが小さい。これは、正規化をしない方が X の中の大きな値に推定が引っ張られ、同じ局所解に落ちやすくなっているからだと考えられる。また、 X の行和1の正規化は物理的には、観測時間内のグループの活動の総和が一定であるという意味をもつ。実際はそうではないので、今回の正規化の仕方はこの実験のみに留める。

NMF の妥当性

NMF, PCA, ICA, logistic regression, glasso の性能の比較を行った。Logistic regression と glasso については筆者の卒論 [42] を参照されたい。どちらも時間窓をスライドさせてネットワークを推定する。今回の実験では時間窓を 40, スライド幅を 20 とした。Glasso のハイパーパラメータを $\rho = 0.3$ とした。一回でもエッジが張られたニューロン同士は同じグループとして推定量 \hat{A} と同じ行列を作成した。

PCA と ICA は NMF と同じく一般化線形成成分分析の手法 [43] である。PCA と ICA では D に相当する行列で NMF と同じように推定量 \hat{A} を作成する。3つの手法の基底数は 10 とした。

1 のタイプについて 100 種類のデータを生成し、 $\hat{A}(X, K)$ を閾値 0.5 で切って $\{0, 1\}^{I \times I}$ の行列にした時の F1 score を比較した。図 4.5 より、NMF の精度が最も高いことが分かる。

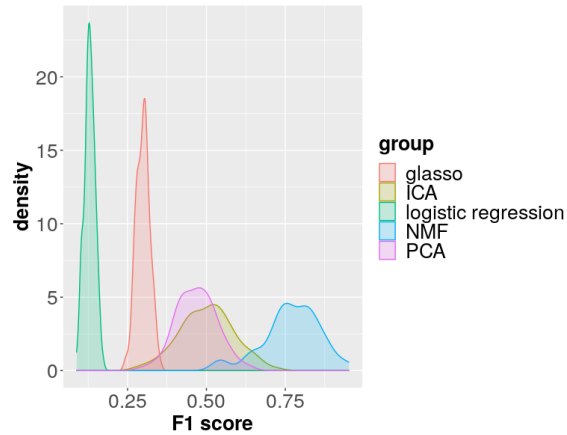


図 4.5: NMF, PCA, ICA, logistic regression, glasso の F1 score の密度分布

NMF の非負制約がデータの生成モデルに合っているためだと思われる。Glasso と logistic regression については精度が低い、各窓ごとのニューロンネットワークの活動を反映している可能性がある、この実験のみで有用性は判断できない。

ブートストラップ法とモデル平均の有用性

ブートストラップ法で $\bar{A}(X, K)$ を推定する有用性を確認するために、タイプ 1 の人工データ 100 個について、1 回 NMF を行った結果、30 回初期値を変えた結果、30 回ブートストラップした結果の F1 score を図 4.6 に示す。F1 score の計算方法は 3.2.2 節と同じである。NMF の基底数は真の基底数 10 を用いた。図 4.6 より、ブートストラップを行った方が精度が高くなるのがわかる。

基底数別に推定された \bar{A} とモデル平均をとった \bar{A} の F1 score を図 4.7 に示す。これより、真の基底数周りの \bar{A} の平均をとることである程度の精度は保たれることがわかる。

NMF の基底数

NMF の基底数を決める方法をいくつか試した。人工データ 86 個について Brunet らと Ubrau らの方法で基底数を決めた時に各基底数が何回選ばれるかを図 4.8, 図 4.9 に示す。Brunet らの方法では初期値を変化させていたが、本実験ではブートストラップの結果で代用している。Brunet らの方法では真の基底数 10 に近い基底数が多く選ばれているが、小さい基底数も選ばれている。Ubaru らの方法では小さい基底数が選ばれる傾向にあった。

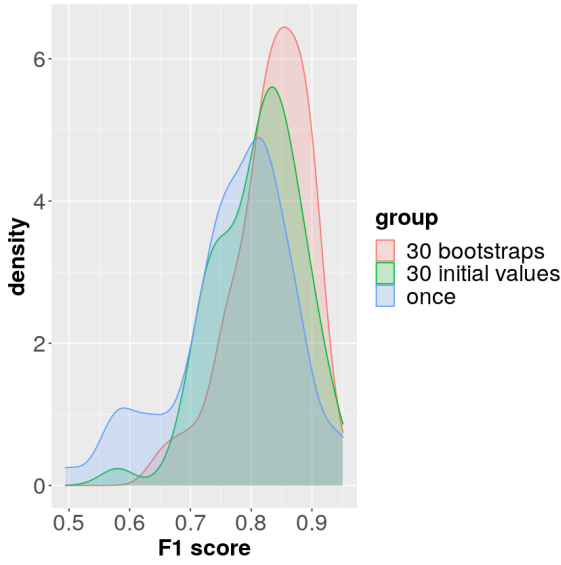


図 4.6: NMF を 1 回行った時の $\hat{A}(X, 10)$, 30 回初期値を変えた $\hat{A}(X, 10)$ の平均, 30 回ブートストラップを行った $\bar{A}(X, K)$; $K = 10$ それぞれの F1 score の分布.

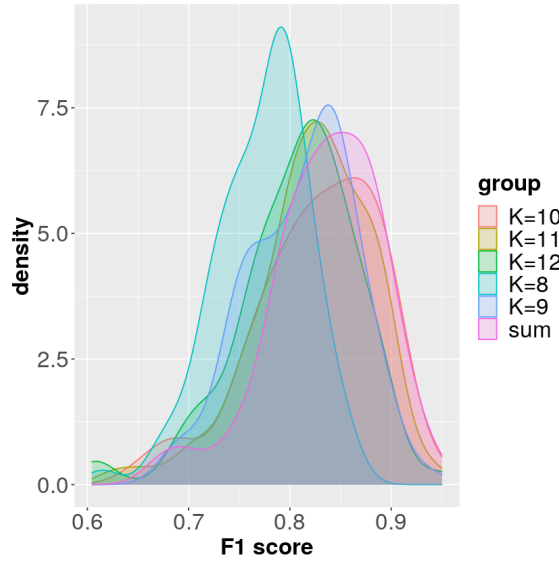


図 4.7: 基底数ごとにブートストラップを行った時の $\bar{A}(X, K)$ の F1 score と全ての基底数についてモデル平均をとった $\bar{A}(X, K)$ の F1 score の分布.

また, 1つの人工データについて AIC と AICc を計算した結果を図 4.10, 図 4.11 に示す. AIC と AICc は最小となるモデルを選択する情報量基準だが, 基底数が大きくなるごとに減少する傾向があった. 他の人工データについても同様の傾向にあった.

4.1.3 \bar{A} のクラスタリングに関する実験

得られた類似度 \bar{A} についてスペクトラルクラスタリングを行い, 精度を確かめる. 比較手法には [13] で性能の良かった ICA-CS を用いる. また, 相関行列を k 近傍法 ($k = 20$) で $\{0, 1\}$ の行列にしたものとも比較を行う. ここで注意したいのは, 実データを扱う際には k をどう設定すればいいかわからない. また, グループ内のニューロン数に大きな差がある場合は k 近傍法ではうまく推定できないことが考えられる. よって比較のために相関行列に易しい設定をとった.

実験設定

人工データのシミュレーション時間を実データと同じになるように 905[s] とする. ただし, 最初の 5[s] は安定性のため解析には用いない. また, 興奮性ニューロンに対する強い外部入力のを平均を 0.5 とした. 作成した人工データの種類を表 4.6 に示す. 900[s] のシミュレーションではニューロングループは 5[s] 活動する機会が 180 回与えられる. 7200[s] のシミュレーションでは 180 回活動する機会がある. all-active は 180 回のうち全てでどれかのグループが活動するデータである. non-active1 は 180 回のうち 120 回活動, non-active2 は 60 回活動するデータである. それぞれのデータを 50 個ずつシミュレーションを行った.

NMF の設定は 3.2 節と同じである. 基底数は 8 から 12 までとして, それぞれ 30 回ブートストラップを行った結果を用いる.

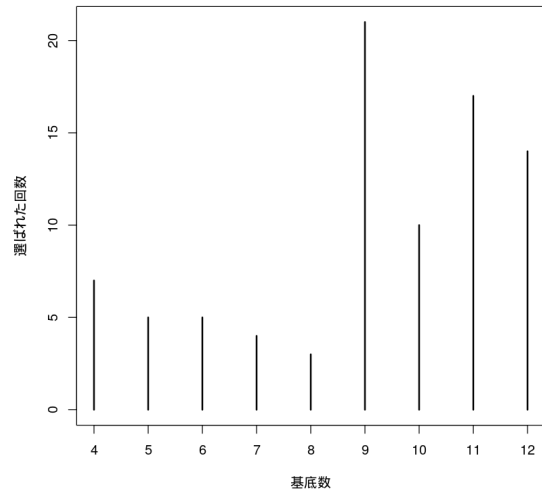


図 4.8: Brunet らの方法で基底数を決めた時に各基底数が選ばれた回数 (真の基底数は 10).

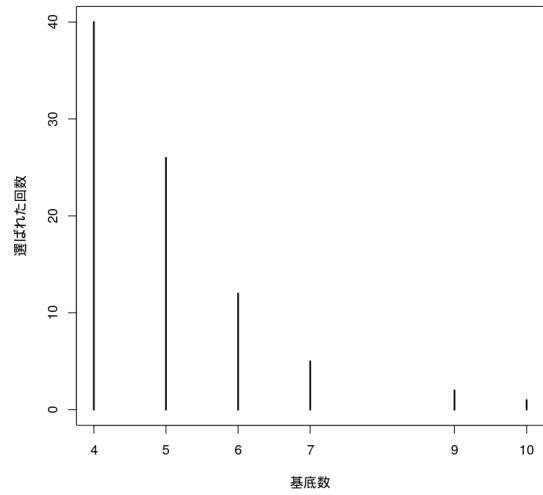


図 4.9: Ubaru らの方法で基底数を決めた時に各基底数が選ばれた回数 (真の基底数は 10).

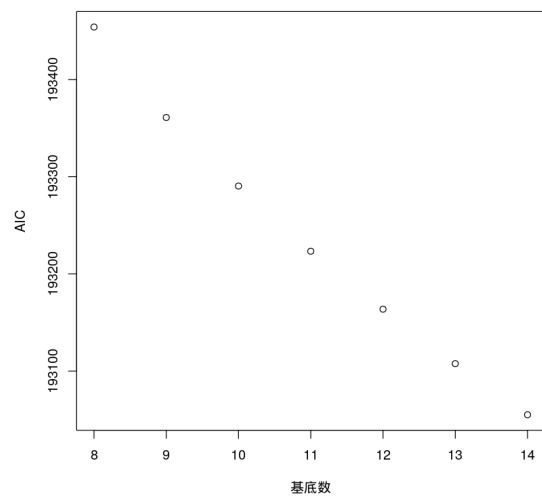


図 4.10: あるデータについて AIC を計算した時の結果.

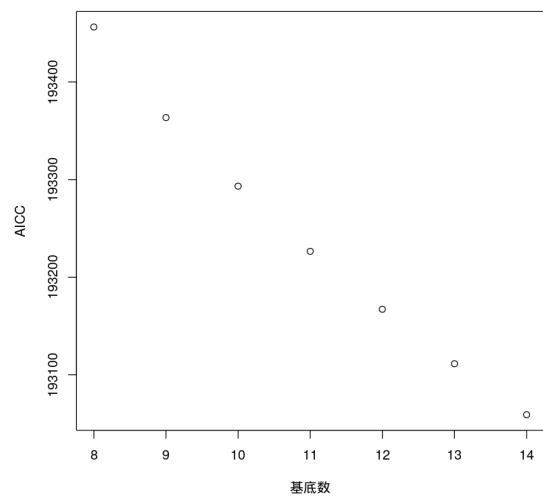


図 4.11: あるデータについて AICc[44] を計算した時の結果.

人工データの種類	特徴
all-active	全活動可能回数で活動
non-active1	活動可能回数 180 回のうち 120 回活動
non-active2	活動可能回数 180 回のうち 60 回活動

表 4.6: 人工データの種類

全てのニューロンが1つのグループに所属する場合

3 種類のデータについて、 \bar{A} と相関行列のスペクトラルクラスタリング、ICA-CS でクラスタリングを行った。真のクラスタとの Best Match score（以降 BMS）を算出した結果を図 4.12 に示す。ニューロングループの活動が多い時は相関行列の BMS が最も高いが、活動

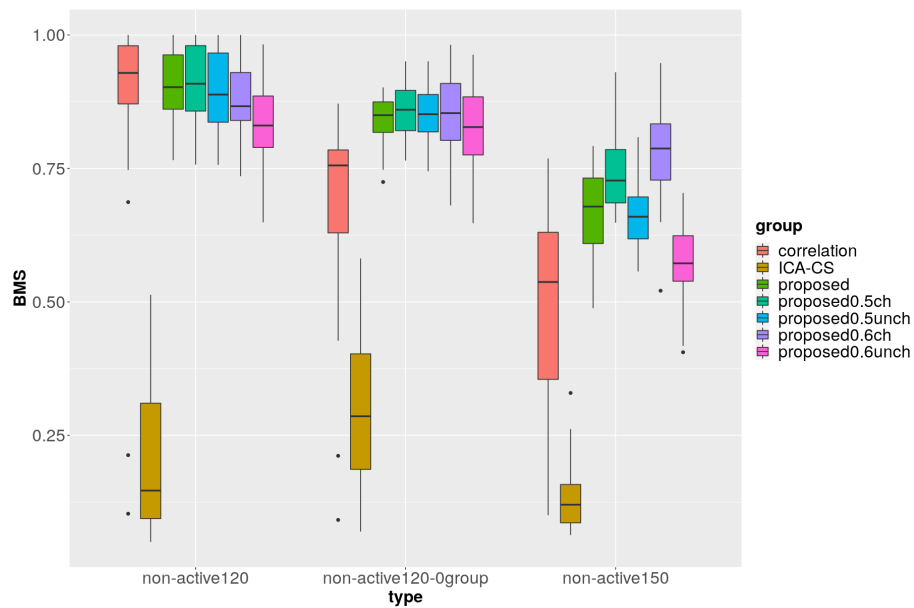


図 4.12: 3 種類の人工データについて提案手法、相関行列、ICA-CS でクラスタリングをした結果。

しない時間が長くなると提案手法の BMS の方が良くなる。相関行列のうち低い BMS があるのは、クラスタ数が適切に決められなかったためと思われる。ICA-CS の BMS は全てのデータにおいて最も低い。元の論文では、ICA-CS のパフォーマンスが安定して高いのはシミュレーション時間が 1800[s] 以上となる時だったため、データ不足が考えられる。

グループに所属しないニューロンがある場合

グループに所属しないニューロンは

4.2 実データ解析

本章では実データを解析する。本章では「NMF1 回の結果」は「初期値を 100 回変えて NMF を行い目的関数が最小となった結果」とする。

4.2.1 実データ

本研究で用いるデータは筑波大学の柳沢研究所で計測されたカルシウムイメージングデータである。本データは、2光子多細胞カルシウムイメージングによって1匹のマウスの大脳皮質1次運動野第2・3層のニューロンのイメージング画像を得た後、人手でROIがつけられ、ニューロンごとの数値データに直されたものである。1時間おきに15分間のイメージングが計6回行われ、各イメージングのサンプリングレートは8[Hz]である。このサンプリングレートは使用された機器の最大のものである。用いられた蛍光タンパク質はGCaMP6sである。実験系は図4.13(A)の系で行われた[45]。観測されたニューロン数は154であった。時間方向には4[s]ごとのマウスの状態(wake, REM, NREM)のラベルが付いており、ニューロンごとに興奮性か抑制性のラベルがついている。図4.2.1にニューロン3つの時系列と時系列方向についているラベル(wake, NREM, REM)を示す。

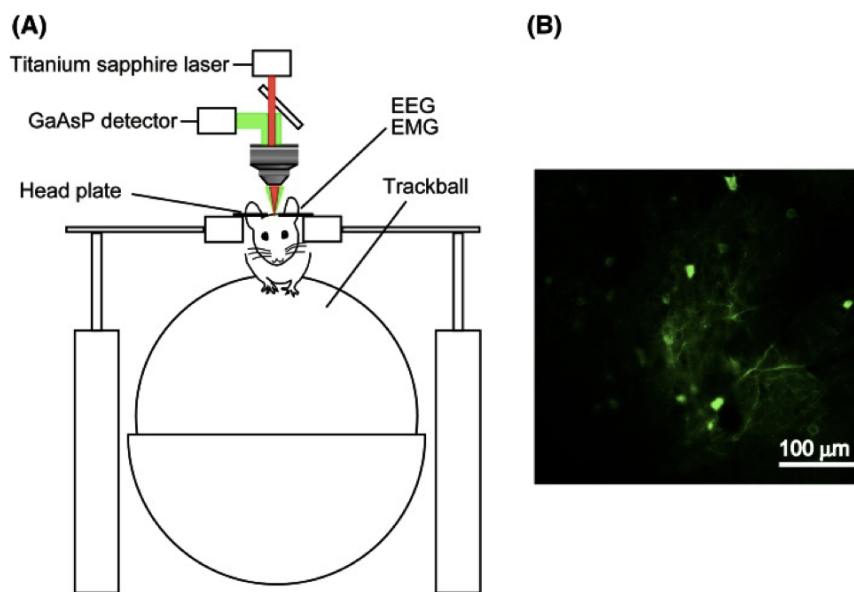


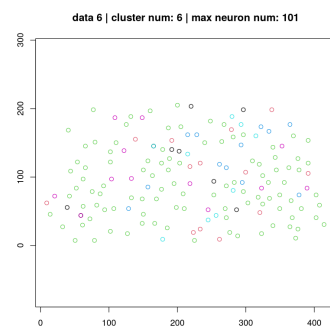
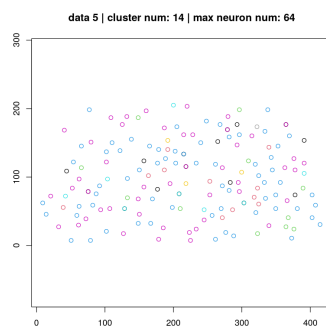
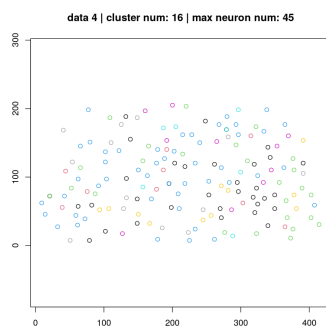
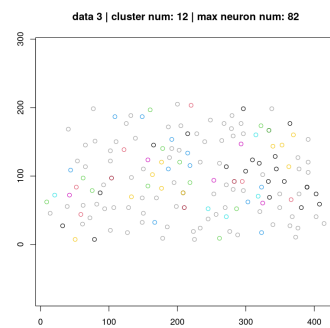
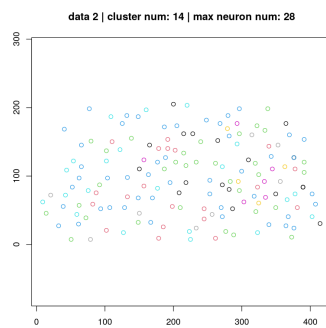
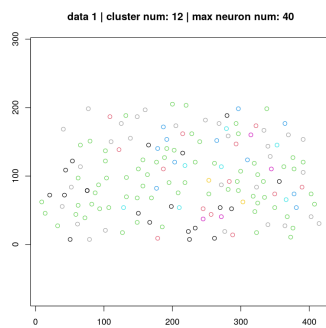
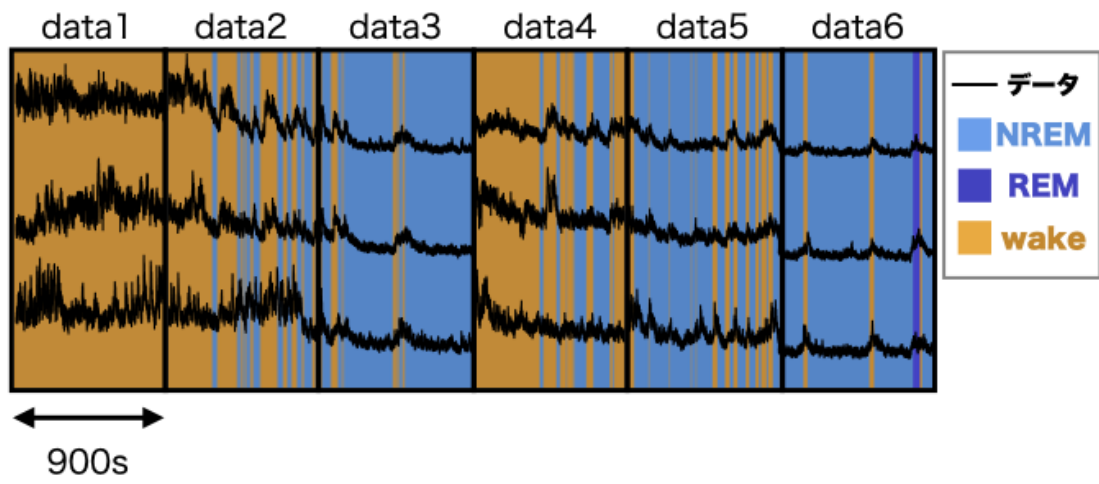
図 4.13: (A) カルシウムイメージングの測定系。マウスは頭を固定されたままボールの上で活動することができる。(B) カルシウムイメージング画像。

4.2.2 実験設定

6回計測されたデータそれぞれについてクラスタリングを行う。NMFを一回行、それを最尤推定の結果とする。最尤推定結果からBICを計算し、目視で基底数の範囲を決める。範囲内の基底数について100回ブートストラップを行い、 \bar{A} を求める。固有値ギャップを算出し、スペクトラルクラスタリングを行う。

4.2.3 結果

推定されたクラスタを図4.2.3に示す。



第5章 結論

本論文では、サンプリングレートの高いカルシウムイメージングデータからニューロンのクラスタリングを行うことを目的に、ブートストラップ法と NMF を用いた隣接行列推定がクラスタリングに有効であることを確かめた。データの生成モデルを立てた上で、ニューロン同士の結合推定に NMF を用い、ブートストラップ法をによって結合確率を推定した。NMF では基底数を決めることが難しいが、モデル平均によってその問題を緩和した。提案アプローチが有効であることを人工データ実験によって確かめた。

今後の展望としては、ニューロンが2つのグループに所属する場合に fuzzy clustering などに行えるかを確かめる必要がある。また、第5章で述べた方法で NMF の真の自由度が測れるかどうか実験を通して確かめたい面白い問題である。

謝辞

本研究を進めるにあたり指導教員の村田昇先生には多くのご指導をいただきました。また、今回用いた技術に関する問題の面白さを知ることができました。深く感謝いたします。赤穂先生、日野先生、有竹さんにも貴重なお時間を割いて助言をいただき、質問に答えていただきました。深く感謝いたします。また、貴重なデータを提供してくださり、カルシウムイメージングやニューロンに関する知見について教えてくださった筑波大学の柳沢研究室の上田助教に感謝いたします。

研究室の先輩、同期、後輩にはゼミなどを通じて多くのアドバイスをいただきました。特に同期の皆様には細かい相談や精神面でも多く支えていただきました。ありがとうございました。

参考文献

- [1] E. M. Izhikevich, J. A. Gally, and G. M. Edelman, “Cerebral Cortex V 14 N 8 Spike-timing Dynamics of Neuronal Groups”, *Cortex August*, vol. 14, pp. 933–944, 2004.
- [2] O. Sporns, “Brain connectivity”, *Scholarpedia*, vol. 2, no. 10, p. 4695, Oct. 2007.
- [3] K. D. Harris, *Cell Assemblies of the Superficial Cortex*, Oct. 2012.
- [4] T. J. Sejnowski, P. S. Churchland, and J. A. Movshon, “Putting big data to good use in neuroscience.”, *Nature neuroscience*, vol. 17, no. 11, pp. 1440–1, Nov. 2014.
- [5] 中村 健, “神経細胞内局所的カルシウム濃度変化のリアルタイムイメージング法”, *Folia Pharmacol. Jpn*, vol. 121, no. 5, pp. 357–364, 2003.
- [6] G. Q. Bi and M. M. Poo, “Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type.”, *The Journal of neuroscience: the official journal of the Society for Neuroscience*, vol. 18, no. 24, pp. 10 464–10 472, 1998.
- [7] 平 理一郎, “脳神経計算原理の解明を目指した 2 光子多細胞イメージングの情報技術展開 [II・完] ”, *Animal Genetics*, vol. 101, no. 9, pp. 926–931, 2018.
- [8] K. Li, L. Guo, J. Nie, G. Li, and T. Liu, “Review of methods for functional brain connectivity detection using fMRI”, *Computerized Medical Imaging and Graphics*, vol. 33, no. 2, pp. 131–139, Mar. 2009.
- [9] J. T. Vogelstein, B. O. Watson, A. M. Packer, R. Yuste, B. Jodynak, and L. Paninski, “Spike Inference from Calcium Imaging Using Sequential Monte Carlo Methods”, *Biophysical Journal*, vol. 97, no. 2, pp. 636–655, Jul. 2009.
- [10] Y. Mishchenko, J. T. Vogelstein, and L. Paninski, “A Bayesian approach for inferring neuronal connectivity from calcium fluorescent imaging data”, *The Annals of Applied Statistics*, vol. 5, no. 2B, pp. 1229–1261, 2011.
- [11] O. Stetter, D. Battaglia, J. Soriano, and T. Geisel, “Model-Free Reconstruction of Excitatory Neuronal Connectivity from Calcium Imaging Signals”, *PLoS Comput Biol*, vol. 8, no. 8, p. 1 002 653, 2012.
- [12] Y. Ikegaya, G. Aaron, R. Cossart, D. Aronov, I. Lampl, D. Ferster, and R. Yuste, “Synfire chains and cortical songs: temporal modules of cortical activity.”, *Science (New York, N.Y.)*, vol. 304, no. 5670, pp. 559–64, Apr. 2004.
- [13] J. Mölter, L. Avitan, and G. J. Goodhill, “Detecting neural assemblies in calcium imaging data”, *BMC Biology*, vol. 16, no. 1, p. 143, Nov. 2018.
- [14] K. Ghandour, N. Ohkawa, C. C. A. Fung, H. Asai, Y. Saitoh, T. Takekawa, R. Okubo-Suzuki, S. Soya, H. Nishizono, M. Matsuo, M. Osanai, M. Sato, M. Ohkura, J. Nakai, Y. Hayashi, T. Sakurai, T. Kitamura, T. Fukai, and K. Inokuchi, “Orchestrated ensemble activities constitute a hippocampal memory engram”, *Nature Communications*, vol. 10, no. 1, pp. 1–14, Dec. 2019.

-
- [15] L. Sjulson and G. Miesenböck, “Optical Recording of Action Potentials and Other Discrete Physiological Events: A Perspective from Signal Detection Theory”, *Physiology*, vol. 22, no. 1, pp. 47–55, Feb. 2007.
 - [16] D. D. Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization”, *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.
 - [17] T. Boongoen and N. Iam-On, *Cluster ensembles: A survey of approaches with recent extensions and applications*, May 2018.
 - [18] L. Breiman, “Bagging Predictors”, Tech. Rep., 1996, pp. 123–140.
 - [19] B. Efron, “Bootstrap Methods: Another Look at the Jackknife”, *The Annals of Statistics*, vol. 7, no. 1, pp. 1–26, Jan. 1979.
 - [20] I. Wasserman, “Bayesian model selection and model averaging”, *Journal of Mathematical Psychology*, vol. 44, no. 1, pp. 92–107, Mar. 2000.
 - [21] H. Akaike, “A New Look at the Statistical Model Identification”, *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.
 - [22] J. P. Brunet, P. Tamayo, T. R. Golub, and J. P. Mesirov, “Metagenes and molecular pattern discovery using matrix factorization”, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 12, pp. 4164–4169, Mar. 2004.
 - [23] S. Ubaru, K. Wu, and K. E. Bouchard, “UoI-NMF cluster: A robust nonnegative matrix factorization algorithm for improved parts-based decomposition and reconstruction of noisy data”, in *Proceedings - 16th IEEE International Conference on Machine Learning and Applications, ICMLA 2017*, vol. 2017-December, Institute of Electrical and Electronics Engineers Inc., 2017, pp. 241–248.
 - [24] S. Wu, A. Joseph, A. S. Hammonds, S. E. Celniker, B. Yu, and E. Frise, “Stability-driven nonnegative matrix factorization to interpret Spatial gene expression and build local gene networks”, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 16, pp. 4290–4295, 2016.
 - [25] L. N. Hutchins, S. M. Murphy, P. Singh, and J. H. Graber, “Position-dependent motif characterization using non-negative matrix factorization”, *BIOINFORMATICS ORIGINAL PAPER*, vol. 24, no. 23, pp. 2684–2690, 2008.
 - [26] A. T. Cemgil, “Bayesian inference for nonnegative matrix factorisation models”, *Computational Intelligence and Neuroscience*, vol. 2009, 2009.
 - [27] J. Kittler, M. Hatef, R. P. Duin, and J. Matas, “On combining classifiers”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.
 - [28] L. I. Kuncheva and D. P. Vetrov, “Evaluation of stability of k-means cluster ensembles with respect to random initialization”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1798–1808, 2006.
 - [29] X. Fu, K. Huang, N. D. Sidiropoulos, and W. K. Ma, “Nonnegative Matrix Factorization for Signal and Data Analytics: Identifiability, Algorithms, and Applications”, *IEEE Signal Processing Magazine*, vol. 36, no. 2, pp. 59–80, 2019.
 - [30] B. Dong, M. M. Lin, and M. T. Chu, “Nonnegative rank factorization—a heuristic approach via rank reduction”, *Numerical Algorithms*, vol. 65, no. 2, pp. 251–274, Feb. 2014.

-
- [31] A. Ben-Hur, A. Elisseeff, and I. Guyon, “A stability based method for discovering structure in clustered data”, Tech. Rep.
 - [32] R. Tibshirani, G. Walther, and T. Hastie, “Estimating the number of clusters in a data set via the gap statistic”, Tech. Rep.
 - [33] U. Von Luxburg, “A Tutorial on Spectral Clustering”,
 - [34] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks”, *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998.
 - [35] J.-S. Jouhanneau, J. Kremkow, A. L. Dornn, and J. F. Poulet, “In Vivo Monosynaptic Excitatory Transmission between Layer 2 Cortical Pyramidal Neurons”, *Cell Reports*, vol. 13, no. 10, pp. 2098–2106, Dec. 2015.
 - [36] C. Holmgren, T. Harkany, B. Svennenfors, and Y. Zilberter, *Pyramidal cell communication within local networks in layer 2/3 of rat neocortex*, Aug. 2003.
 - [37] E. M. Izhikevich, “Simple Model of Spiking Neurons”, *IEEE TRANSACTIONS ON NEURAL NETWORKS*, vol. 14, no. 6, 2003.
 - [38] S. Song, P. J. Sjöström, M. Reigl, S. Nelson, and D. B. Chklovskii, “Highly non-random features of synaptic connectivity in local cortical circuits”, in *PLoS Biology*, vol. 3, 2005, pp. 0507–0519.
 - [39] B. O. Watson, D. Levenstein, J. P. Greene, J. N. Gelinas, and G. Buzsáki, “Network Homeostasis and State Dynamics of Neocortical Sleep”, *Neuron*, vol. 90, no. 4, pp. 839–852, May 2016.
 - [40] T. W. Chen, T. J. Wardill, Y. Sun, S. R. Pulver, S. L. Renninger, A. Baohan, E. R. Schreiter, R. A. Kerr, M. B. Orger, V. Jayaraman, L. L. Looger, K. Svoboda, and D. S. Kim, “Ultrasensitive fluorescent proteins for imaging neuronal activity”, *Nature*, vol. 499, no. 7458, pp. 295–300, 2013.
 - [41] N. Guan, D. Tao, Z. Luo, and B. Yuan, “NeNMF: An optimal gradient method for nonnegative matrix factorization”, *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2882–2898, Jun. 2012.
 - [42] 永山瑞生, “マウスのカルシウムイメージングデータを用いた睡眠・覚醒状態の解析”, 2018.
 - [43] A. Cichocki, R. Zdunek, A. H. Phan, and S.-i. Amari, *Nonnegative Matrix and Tensor Factorizations*. Chichester: Wiley Publishing, 2009.
 - [44] M. R. Symonds and A. Moussalli, *A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike’s information criterion*, Aug. 2011.
 - [45] T. Kanda, N. Tsujino, E. Kuramoto, Y. Koyama, E. A. Susaki, S. Chikahisa, and H. Funato, “Sleep as a biological problem: an overview of frontiers in sleep research”, *The Journal of Physiological Sciences*, vol. 66, no. 1, pp. 1–13, Jan. 2016.
 - [46] M. Babaei, S. Tsoukalas, M. Babaei, G. Rigoll, and M. Datcu, “Discriminative Nonnegative Matrix Factorization for dimensionality reduction”, *Neurocomputing*, vol. 173, pp. 212–223, Jan. 2016.
 - [47] V. C. Cheung, K. Devarajan, G. Severini, A. Turolla, and P. Bonato, “Decomposing time series data by a non-negative matrix factorization algorithm with temporally constrained coefficients”, in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2015-Novem, Institute of Electrical and Electronics Engineers Inc., Nov. 2015, pp. 3496–3499.
-

- [48] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, “Sparsity and smoothness via the fused lasso”, Tech. Rep., 2005, pp. 91–108.

付録 A 人工データの作成方法

付録B 考案したNMFの制約

本節では、データの特徴を考慮して考案したNMFの制約とモデルエビデンスの計算方法を紹介する。

1 分解能の決定

ニューロンの活動データの扱いには時間分解能と空間分解能の2つの側面から検討する必要がある。時間分解能については、蛍光強度データをそのまま用いる、時間窓に区切るなどが考えられる。空間分解能については、ニューロン1個を見る場合、2個を見る場合、複数を見る場合が考えられる。手法によってどのレベルでデータを扱うかが異なる。表1にカルシウムイメージングデータを解析する際に使えそうな手法を載せる。これらの手法ではニューロンのネットワーク構造変化やグループの活動の変化などを観察できる。提案アプローチでは行列分解を使いニューロングループを推定し、グループの活動時系列の情報は使わなかった。しかし、提案アプローチでも時間窓で区切ってニューロングループの推定を行えばクラスタの変化は抽出できる。その際、あまりに短い時間窓内では意味のある情報が取り出せない可能性がある。ニューロングループがどれくらいの頻度で活動するのかをある程度の仮定を置いて時間窓を決定する必要がある。

	生データ	時間窓で区切る
ペアで見る	時系列クラスタリング, TE	glasso, 類似度+クラスタリング
複数で見る	行列分解	ロジスティック回帰, 時系列クラスタリング

表 1: カルシウムイメージングデータ解析に使えそうな手法

2 スケール除去

ニューロンごとに発現している蛍光タンパク質の量や細胞の大きさが異なる。そのため、ニューロンごとのバイアスが観測データに載っていると考え、バイアスを除去する方法を試した。この時の数理モデルは以下になる：

$$X = DC + H + B,$$

ただし、 $B \in \mathbb{R}_{+}^{I \times J}$ は行ごとに同じ数値が入ったバイアス行列である。

バイアスの推定方法は、 D に1列を足し、 C に1の1行を足してNMFを更新する。 D の列にバイアスが推定されることを期待した。

簡単な人工データ実験を行った結果、足したバイアスよりも大きいバイアスが推定されてしまうことがわかった。また、そもそもの数理モデルが異なると考え直した。

蛍光タンパク質の量や細胞の大きさが異なるということは、各ニューロンはスケールされていると考えられ、以下のように表される：

$$X = A(DC + H),$$

ただし、 $A \in \mathbb{R}_{+}^{I \times I}$ は対角行列である。

ナイーブな求め方は、 A は単位行列で初期化し、NMF 一回の更新ごとに X をニューロンごとに残差の四分位範囲で割り、その値を A にかけていく。簡単な人工データ実験で乗法更新則を用いた場合は A は真の値に近いものが推定された。しかし、第 4 章で作成した人工データ実験で Nesterov 更新を行った際は時々おかしい局所解に落ちてしまった。

3 重複除去

置いた仮定では、あるニューロンが複数のグループに所属する時、グループの活動は被らないとしている。しかし、NMF の推定時にそのような制約は入れていないので、NMF で推定した結果この仮定が破られているようであれば制約は入れなければならない。

以下の目的関数を考えた：

$$\arg \min_{D \geq 0, C \geq 0} \|X - DC\|_F^2 - \lambda \sum_{k=1}^K \sum_{l \neq k}^K (\|d_{:l} - d_{:k}\|_1 \|c_{l:} - c_{k:}\|_1). \quad (1)$$

更新則を導出する。参考にしたのは [46] である。式 (1) の Lagrange 関数 L は、

$$L = \text{Tr}(X^T X) - 2\text{Tr}(X^T DC) + \text{Tr}(C^T D^T DC) - \text{Tr}(\Phi_C C^T) - \text{Tr}(\Phi_D D^T) - \lambda \text{Tr}(F^T C H^T S^T D F),$$

であり、KKT 条件は、

$$\begin{aligned} \frac{\partial L}{\partial C} &= \frac{\partial L}{\partial D} = 0 \\ D &\geq 0 \\ C &\geq 0 \\ \Phi_C &\geq 0 \\ \Phi_D &\geq 0 \\ \Phi_C C &= \Phi_D D = 0 \end{aligned}$$

である。ただし、 Φ_D と Φ_C はそれぞれ $D \geq 0$, $C \geq 0$ に対する Lagrange 乗数で、 $F \in [0, 1]^{K \times (K-1)!}$ は 2 つの時間の組み合わせを表現した以下のような行列である：

$$F = \begin{pmatrix} 1 & 1 & \dots & 0 & \dots & 0 \\ -1 & 0 & \dots & 1 & \dots & 0 \\ 0 & -1 & \dots & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \dots & 1 \\ 0 & 0 & \dots & 0 & \dots & -1 \end{pmatrix}.$$

また、 $S = \text{sign}(DF)$, $H = \text{sign}(F^T C)$ とおく。

D を求めるには以下の式を解く：

$$\frac{\partial L}{\partial D} = -2XC^T + 2DCC^T - \Phi_D - \lambda SHC^T FF^T = 0.$$

これを求めると D の要素の更新は以下である：

$$d_{ik} \leftarrow d_{ik} \frac{2[XC^T]_{ik} + \lambda[SHC^T FF^T]_{ik}^+}{2[DCC^T]_{ik} - \lambda[SHC^T FF^T]_{ik}^-},$$

ただし, $[\cdot]^+$ は行列の中の正の要素, $[\cdot]^-$ は負の要素である.

同様に, C の要素の更新は以下の通りである:

$$c_{kj} \leftarrow c_{kj} \frac{2[D^T X]_{kj} + \lambda[FF^T D^T SH]_{kj}^+}{2[D^T DC]_{kj} - \lambda[FF^T D^T SH]_{kj}^-}.$$

しかし, ハイパーパラメータ λ の最適な決め方は分からない. また, λ が大きいと D と C はスパースになるが, 本来 2 つの基底に所属するニューロンの $d_{i\cdot}$ が 1 つの基底以外 0 に近くなることが実験を通してわかった. この手法は改善の余地がある上, 本当にこの制約が必要かは実際のデータを NMF で分解した際にどれくらい仮定に沿っていないかを確認する必要がある.

4 時間方向への制約

カルシウムイメージングデータはスパイク情報を反映するのが遅く, 一度上がった蛍光強度は緩やかに下がっていく. そのため, NMF で分解を行う際も, 行列 C の時間方向に前時刻の値と近くなるような制約を入れることでより正確なニューロングループの抽出が行えると考えられる. C の偶数列を前後の列の平均とする NMF も提案されている [47]. しかし, これはかなりスムーズになる制約だと考えられる. そこで, 以下のような制約を加えた目的関数が考えられる:

$$\arg \min_{D \geq 0, C \geq 0} \|X - DC\|_F^2 + \lambda \sum_t \|c_{:t} - c_{:t-1}\|_1.$$

これは fused lasso [48] と同じような制約である.

更新則は, D はユークリッド型 NMF と同じだが C は異なる. 更新則は以下である:

$$c_{kj} \leftarrow c_{kj} \frac{2[D^T X]_{kj} - s_{kj}}{2[D^T DC]_{kj}},$$

ただし, $s_{:j} = \text{sign}(c_{:j} - c_{:j-1})$ であり, 1 列目のみ $s_{:1} = \mathbf{0}$ である.

重複除去制約と同じく, λ の決め方が分からないため使用は断念した. NMF は初期値依存の問題があるため, λ を変化させると解空間も変化し, 結果は滑らかに変化しない. そのため, λ を変化させた効果を比較するのは難しい.

付録C NMFのモデルエビデンス

NMFの基底数を決める際にブートストラップから計算されたモデルエビデンスを用いることを考える。BICは最尤推定量の尤度からモデルエビデンスを近似して扱っている。ブートストラップによってモデルエビデンスを近似計算できると考えられる。本節ではその説明をする。

基底数 k のNMFのモデルを \mathcal{M}_k とおく。ノイズ行列 H の各要素が正規分布 $\mathcal{N}(0, \sigma^2)$ に従う \mathcal{M}_k の尤度は以下である。

$$p(X|Y_k, \mathcal{M}_k) = \prod_{i,j} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{([Y_k]_{ij} - x_{ij})^2}{2\sigma^2}\right),$$

ただし、 $Y_k = D_k C_k$ で、モデル \mathcal{M}_k における推定量である。対数尤度は以下ようになる。

$$\log p(X|Y_k, \mathcal{M}_k) = -\frac{IJ}{2}(\log 2\pi + 2\log \sigma) - \frac{1}{2\sigma^2} \sum_{ij} ([Y_k]_{ij} - x_{ij})^2.$$

今回の問題を Bayesian model averaging の枠組みに当てはめると、

$$p(A|X) = \sum_k p(A|\mathcal{M}_k, X) p(\mathcal{M}_k|X)$$

で $p(A|X)$ を求めることになる。今回は異なる基底数のNMFから求めた A をモデルの事後確率で重み付けて足し合わせることになる。

モデルの事後確率は

$$p(\mathcal{M}_k|X) = \frac{m_k p(\mathcal{M}_k)}{\sum_l m_l p(\mathcal{M}_l)}$$

で表される。ただし、

$$m_k = \int p(X|Y_k, \mathcal{M}_k) p(Y_k|\mathcal{M}_k) dY_k$$

である。これはモデルエビデンスや marginal likelihood と呼ばれる。また、 $p(\mathcal{M}_k)$ はモデルが正しい確率である。

ある条件下で、母数の事後確率の密度関数はブートストラップによる最尤推定量の分布と同じになる。そのため、ブートストラップサンプルから推定した A の分布をもとに事後確率を計算すれば良い。

現在の設定では $p(\mathcal{M}_k)$ は一様なため、エビデンスを見る必要がある。

0.1 ラプラス近似

エビデンスの計算には事前分布を決めなければならず、積分も計算しなければいけない。そこでラプラス近似により m_k は以下のような \hat{m}_k で近似できる。

$$\log \hat{m}_k = \log p(X|\hat{Y}_k, \mathcal{M}_k) - \frac{d_k}{2} \log n$$

ここで, d_k は母数の数 ($I \times K + K \times J$), n は観測データ数 (J) である. この近似を使って Bayes 因子のログを計算したのが BIC である. Bayes 因子は 2 つのモデルエビデンスの比である.

$\log p(X|\hat{Y}_k, \mathcal{M}_k)$ は初期値を変えて NMF を行い, 尤度が最も大きくなった対数尤度を用いる. また, $\sigma^2 = \text{Var}(X - Y_k)$ として計算する.

0.2 ブートストラップによる近似

ブートストラップの推定量の分布は最尤推定量の分布を近似する (ブートストラップサンプルが生成されたパラメータ分バイアスは乗る) ので, m_k は以下のように近似できる.

$$m_k = \int p(X|Y_k, \mathcal{M}_k)p(Y_k|\mathcal{M}_k)dY_k \\ \sim \frac{1}{B} \sum_b \prod_{i,j} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{([Y_k^b]_{ij} - X_{ij}^b)^2}{2\sigma^2}\right)$$

ここで, Y_k^b はブートストラップサンプル b から計算された Y_k で, B はブートストラップサンプル数である. また, $\sigma^2 = \text{Var}(X^b - Y_k^b)$ として計算する.

0.3 実験

人工データセット 86 個について 2 つの近似方法でモデルエビデンスを計算した. その結果を図 1 と図 2 に示す. 同じデータの結果を線で結んでいる. ラプラス近似を行った場合は真の基底数 10 付近で最大値をとっている. ブートストラップによる近似を行った場合は基底数に比例してエビデンスの値も大きくなっている. また, 同じデータでも基底数によってオーダーが大きく異なるが, これは尤度計算に用いる分散をデータから計算しているためである.

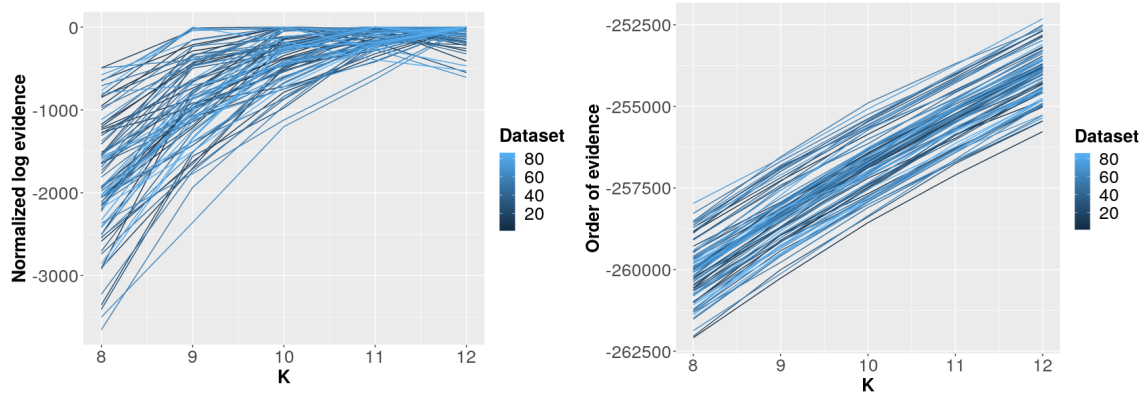


図 1: ラプラス近似によって計算されたログエビデンス. 各データについて最大値が 0 となるように定数を足した.

図 2: ブートストラップ近似によって計算されたエビデンスのオーダー (10 を底とした指数).

また, 図 2 のログを取り, 図 1 とプロットした図を図 3 に示す. これより, ブートストラップによる近似を用いた場合のエビデンスの値は大きいことが分かる. ラプラス近似では $\frac{d_k}{2} \log n$ によって補正された値である. NMF ではデータが増えると補正の値も増えるためラプラス近似は適さないが, 図 3 の 2 つの近似手法のエビデンスの差が過剰に足された補正の値だとも考えることができる.

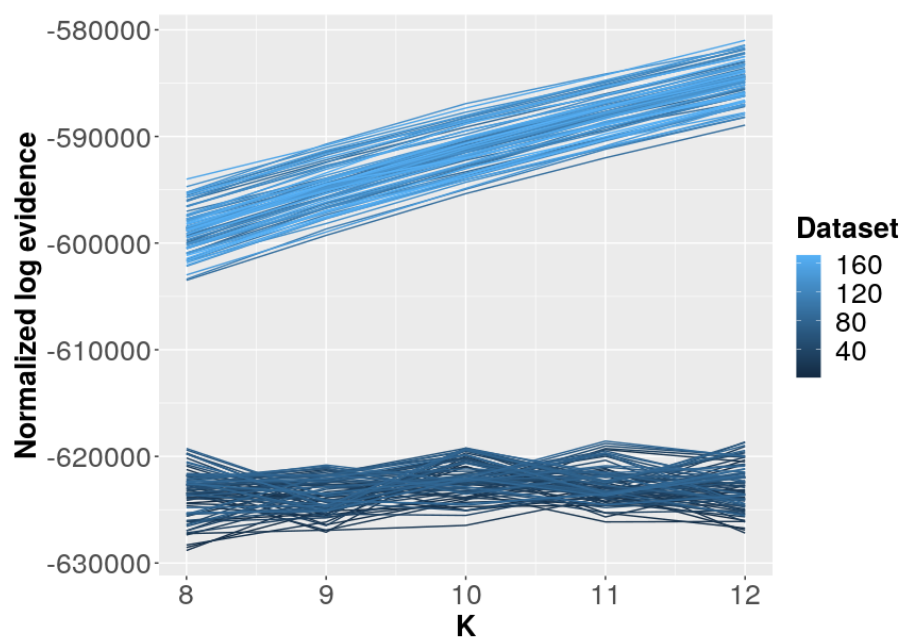


図 3: 2つの近似方法によって計算されたログエビデンス.