Tracking Neural Activity

Automated Image Analysis

B. Jones\textsuperscript{1} A. Stanley\textsuperscript{2}

\textsuperscript{1}Department of Mathematics
State University of New York at Geneseo

\textsuperscript{2}Department of Mathematics
Grand Valley State University

January 26, 2019
Outline

1. Our Motivation

2. Methods of Automated Cell Finding
   - Dimension Reduction
   - Singular Value Decomposition

3. Aligning Sessions

4. Eliminating Algorithmic Error
   - Improving Cell Substructure Modeling
Outline

1. Our Motivation

2. Methods of Automated Cell Finding
   - Dimension Reduction
   - Singular Value Decomposition

3. Aligning Sessions

4. Eliminating Algorithmic Error
   - Improving Cell Substructure Modeling
Understanding the Brain

- We wish to understand how brain activity correlates with specific actions
- Current leading algorithms to detect cells need improvement
  - Based off of generic criteria (size and roundness) - not ideal
  - Not all cells are perfectly round
  - Not all cells are perfectly visible in the brain images
  - There is also background noise, or neuropil
  - Only detect $\sim50\%$ of cells, leading scientists to manually pick out cells
The Zaneta Mouse Data Set: An Example
Outline

1. Our Motivation

2. Methods of Automated Cell Finding
   - Dimension Reduction
   - Singular Value Decomposition

3. Aligning Sessions

4. Eliminating Algorithmic Error
   - Improving Cell Substructure Modeling
This is Your Brain on Math

We represent a data set (a video of the brain) as a 3-dimensional matrix $\mathbf{Y} \in \mathbb{R}^{N \times D_y \times D_x}$ where

\[ N = \text{number of frames} \]
\[ D_y = \text{number of pixels along } y \]
\[ D_x = \text{number of pixels along } x. \]
This is Your Brain on Math

We represent a data set (a video of the brain) as a 3-dimensional matrix \( \mathbf{Y} \in \mathbb{R}^{N \times D_y \times D_x} \) where

\[
\begin{align*}
N &= \text{number of frames} \\
D_y &= \text{number of pixels along y} \\
D_x &= \text{number of pixels along x}.
\end{align*}
\]

Note that \( \mathbf{Y} \) is very large.

In order to make these data sets more manageable, we convert to a rectangular array with dimensions \( N \times D \) where \( D = D_x \times D_y \)
From 3-Dimensional to 2-Dimensional

We reshape a $4 \times 5 \times 4$ matrix into a single $4 \times 20$ matrix.
From 3-Dimensional to 2-Dimensional

We reshape a $4 \times 5 \times 4$ matrix into a single $4 \times 20$ matrix.

**QUESTION:**
How do we work with our high-dimensional $N \times D$ array and find our sparse set of sources (neurons)?
Dimension Reduction

Each source has a signal and a footprint. A neuron’s **signal** is a measurement over time (activity). A neuron’s **footprint** is its spacial measurement (location, size, and shape).

We aim to find these true sources (neurons) in our data and eliminate background noise caused by neuropil.
Leading Algorithms

To identify potential sources, or “regions of interest” (ROIs), Suite2P uses singular value decomposition (SVD) and CalmAn uses non-negative matrix factorization (NNMF).
Synthetic Neurons

To understand how these programs use SVD, we randomly-generated a mini video $A \in \mathbb{R}^{151 \times 300 \times 300}$ of synthetic neurons including...
Synthetic Neurons

To understand how these programs use SVD, we randomly-generated a mini video $A \in \mathbb{R}^{151 \times 300 \times 300}$ of synthetic neurons including

- Individual signals
- Individual footprints
- Background noise
Synthetic Neurons

To understand how these programs use SVD, we randomly-generated a mini video $A \in \mathbb{R}^{151\times 300\times 300}$ of synthetic neurons including

- Individual signals
- Individual footprints
- Background noise

Note that we generated 15 neurons in $A$. 
Singular Value Decomposition

Let \( A \) be an \( m \times n \) matrix.

\[
A = U \Sigma V^T
\]

where

- \( U \) is an \( m \times m \) orthogonal matrix
- \( \Sigma \) is an \( m \times n \) diagonal matrix of the singular values \( \sigma_1, \sigma_2, \ldots, \sigma_r \) of \( A \) where \( r = \min(m, n) \)
- \( V \) is an \( n \times n \) orthogonal matrix
Singular Value Decomposition

Let $A$ be an $m \times n$ matrix.

$$A = U \Sigma V^T$$

where

- $U$ is an $m \times m$ orthogonal matrix
- $\Sigma$ is an $m \times n$ diagonal matrix of the singular values $\sigma_1, \sigma_2, \ldots, \sigma_r$ of $A$ where $r = \min(m, n)$
- $V$ is an $n \times n$ orthogonal matrix

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix}_{4 \times 3} = \begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ u_{21} & u_{22} & u_{23} & u_{24} \\ u_{31} & u_{32} & u_{33} & u_{34} \\ u_{41} & u_{42} & u_{43} & u_{44} \end{bmatrix}_{4 \times 4} \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix}_{3 \times 3}$$
Singular Value Decomposition

Let $A$ be an $m \times n$ matrix.

$$A = U \Sigma V^T$$

where

- $U$ is an $m \times m$ orthogonal matrix
- $\Sigma$ is an $m \times n$ diagonal matrix of the singular values $\sigma_1, \sigma_2, \ldots, \sigma_r$ of $A$ where $r = \min(m, n)$
- $V$ is an $n \times n$ orthogonal matrix

Let $u_i$ and $v_i$ be the column vectors of $U$ and $V$ respectively. Then

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$
**Synthetic Neurons**

The largest singular values in $\Sigma$ will contribute most to $A$. Since our goal is to find neurons, the factors that contribute most to our data set, these values become very important.
Synthetic Neurons

The largest singular values in $\Sigma$ will contribute most to $A$. Since our goal is to find neurons, the factors that contribute most to our data set, these values become very important.

After about the fifth singular value, the values become insignificant.
Synthetic Neurons

Let $k = 5 \leq r$. 
Synthetic Neurons

Let $k = 5 \leq r$. Recall

$$A = \sum_{i}^{r} \sigma_{i} u_{i} v_{i}^{T}$$

$$= \sigma_{1} u_{1} v_{1}^{T} + \sigma_{2} u_{2} v_{2}^{T} + \ldots + \sigma_{r} u_{r} v_{r}^{T}$$

$$= \sigma_{1} u_{1} v_{1}^{T} + \sigma_{2} u_{2} v_{2}^{T} + \ldots + \sigma_{k} u_{k} v_{k}^{T} + \ldots + \sigma_{r} u_{r} v_{r}^{T}$$
Synthetic Neurons

Let $k = 5 \leq r$. Recall

\[
A = \sum_{i}^{r} \sigma_i u_i v_i^T
\]

\[
= \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \ldots + \sigma_r u_r v_r^T
\]

We can approximate $A$ using $k$:

\[
\tilde{A} = \sum_{i}^{k} \sigma_i u_i v_i^T
\]

\[
= \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \ldots + \sigma_k u_k v_k^T
\]
Synthetic Neurons

Using our approximation $A$, we potentially have a clearer view of the neurons in our video, eliminating the background noise:
Outline

1. Our Motivation

2. Methods of Automated Cell Finding
   - Dimension Reduction
   - Singular Value Decomposition

3. Aligning Sessions

4. Eliminating Algorithmic Error
   - Improving Cell Substructure Modeling
Dun Mao Data

The goal of the lab at Lethbridge is to study how memory and learning affects the brain.

Matching cells between the sessions will allow us to see, as the mouse repeats actions, how neural firing activity changes:

- Location
  - Which area of the brain is active
  - Which/how many cells in that area are active
- Frequency
- Coordinated firing
Mean Image
ROI Transformation

Trace of ROI 148 from session plane3-Nk650 and trace of ROI 82 from session plane3-Nk650
ROI Transformation
ROI Transformation
ROI Transformation
ROI Transformation
Aligning Sessions

ROI Transformation

The two sessions are aligned (left: 9/26 at 170µm will become green/blue stars, middle: 10/1 at 180µm will become purple/red crosses)
Matching Cells
Matching Cells
Matching Cells

Here, we see that three of the sixteen aligned sessions did not match any cells. This was most likely due to the lack of visual geometric features.
Outline

1. Our Motivation

2. Methods of Automated Cell Finding
   - Dimension Reduction
   - Singular Value Decomposition

3. Aligning Sessions

4. Eliminating Algorithmic Error
   - Improving Cell Substructure Modeling
Cell Substructure
Cell Substructure

Frame: 173

Frame: 676
Principal Component Analysis

Definition

The **principal components** of a data set are a set of linearly uncorrelated variables that account for the variability in the data. The first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and the remaining principal components continue to decrease or stay the same in the amount of variability they account for.
Principal Component Analysis

Definition

The **principal components** of a data set are a set of linearly uncorrelated variables that account for the variability in the data. The first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and the remaining principal components continue to decrease or stay the same in the amount of variability they account for.

Leading algorithms assume there is one principal component contributing to the majority of cell activity (the other components are background noise).
Principal Component Analysis

![Principal Components](image)

```
PC 1
```

```
PC 2
```

```
PC 3
```
Summary

To analyze videos of mice brains, we used

- Singular Value Decomposition

in order to

- Match neurons from different recording sessions to help study memory and learning
- Establish that future algorithms should allow for multiple principal components in each cell.
Acknowledgements

Many thanks to the other students in our lab and our sponsors and mentors at Michigan State University, Drs. Mark Reimers, Michael Moore, and Samanthule Nola.

Thanks also to all the volunteers and sponsors at NCUWM for giving us the opportunity to talk!

And thank you for listening!